



Genética cuantitativa para mejoradores de plantas

Guillermo H. Eyhérbide

INTA Ediciones

Colección
EDUCACIÓN SUPERIOR

Genética cuantitativa para mejoradores de plantas

Ph.D. Guillermo Hugo Eyhérabide



Ministerio de Agricultura,
Ganadería y Pesca
Argentina

INTA Ediciones
DNA de Comunicación Institucional
2022

575 Eyhérabide, Guillermo Hugo
Ey39 Genética cuantitativa para mejoradores de plantas / Guillermo Hugo Eyhérabide. –
Buenos Aires : Ediciones INTA, DNA de Comunicación Institucional , 2022.
481 p. : il. (en PDF)

ISBN 978-987-679-336-0 (digital)

i. título

GENETICA – GENETICA CUANTITATIVA – GENETICA DE POBLACIONES – MODELOS LINEALES –
HETEROSIS – PLANTAS

DD-INTA

Este documento es resultado del financiamiento otorgado por el Estado Nacional,
por lo tanto, queda sujeto al cumplimiento de la Ley N° 26.899.

Diseño: Área de Comunicación Visual - Gerencia de Multimedia.

Este libro cuenta
con licencia:



CONTENIDO

PRÓLOGO 13

CAPÍTULO 1 15

Genética cuantitativa y variación continua

1.1. La variación continua	15
1.2. La Genética Cuantitativa	16
1.3. Reseña histórica de la Genética Cuantitativa	16
1.4. Tipos de caracteres	19
1.4.1. Caracteres cualitativos, mendelianos o discretos	19
1.4.2. Caracteres cuantitativos, métricos, continuos o poligénicos	19
1.5. Arquitectura de los caracteres cuantitativos	23

CAPÍTULO 2 25

Genética de poblaciones

2.1. Población genética	25
2.2. Poblaciones genéticas y equilibrio de Hardy-Weinberg	26
2.3. Número o tamaño efectivo poblacional y deriva genética	28
2.3.1. Cálculo del tamaño o número efectivo	29
2.3.1.1. Efecto del hábito reproductivo	29
2.3.1.2. Efecto de la variabilidad en el número de descendientes por individuo	30
2.3.1.3. Efecto de la variabilidad en el tamaño efectivo poblacional a través de generaciones	31
2.4. Absorción de alelos y deriva genética	32
2.5. Número efectivo y probabilidad de fijación por selección del alelo favorable	33
2.6. Desequilibrio de ligamiento y desequilibrio gamético	33
2.7. Variabilidad genética	39



2.7.1. Clases genotípicas y número de genotipos cuando hay dos alelos por locus	39
2.7.2. Número de clases genotípicas posibles cuando hay múltiples alelos por locus	41
2.7.3. Número de clases fenotípicas	43
2.7.4. Proporción de individuos heterocigotas y homocigotas	43
2.8. Consideraciones sobre muestreo poblacional	
Probabilidad de encontrar un genotipo deseado	48
2.8.1. Frecuencia de individuos con el genotipo óptimo en una población en EHW	51
2.9. Determinación del tamaño de muestra adecuado en función de la frecuencia de individuos con el genotipo deseado	54

CAPÍTULO 3.....57
Modelos genéticos

3.1. Modelo genético básico	57
3.2. Interacción epistática	64
3.3. Epistasis estadística	68
3.4. Medias generacionales	68
3.4.1. Medias generacionales bajo sistemas de endocria (autofecundación y retrocruzamiento)	69
3.4.2. Medias generacionales derivadas del cruzamiento entre padres homocigotas contrastantes bajo apareamiento aleatorio	71
3.4.3. Expresión de una media generacional como función lineal de otras generaciones	72
3.4.4. Medias generacionales y epistasis	75
3.4.5. Media poblacional bajo condición de equilibrio de Hardy-Weinberg	76
3.4.6. Media de poblaciones endocriadas	79
3.4.6.1. <i>Endogamia y media poblacional</i>	81
3.5. Efectos medios de los alelos y efecto de sustitución de un gen	82
3.6. Valor de mejora de un individuo	86
3.7. Predicción genómica del valor de mejora de un individuo	89
3.8. Media poblacional y epistasis	93
3.9. Medias de familias extraídas de poblacionales en EHW	95

CAPÍTULO 4.....101
Media de cruzamientos y heterosis

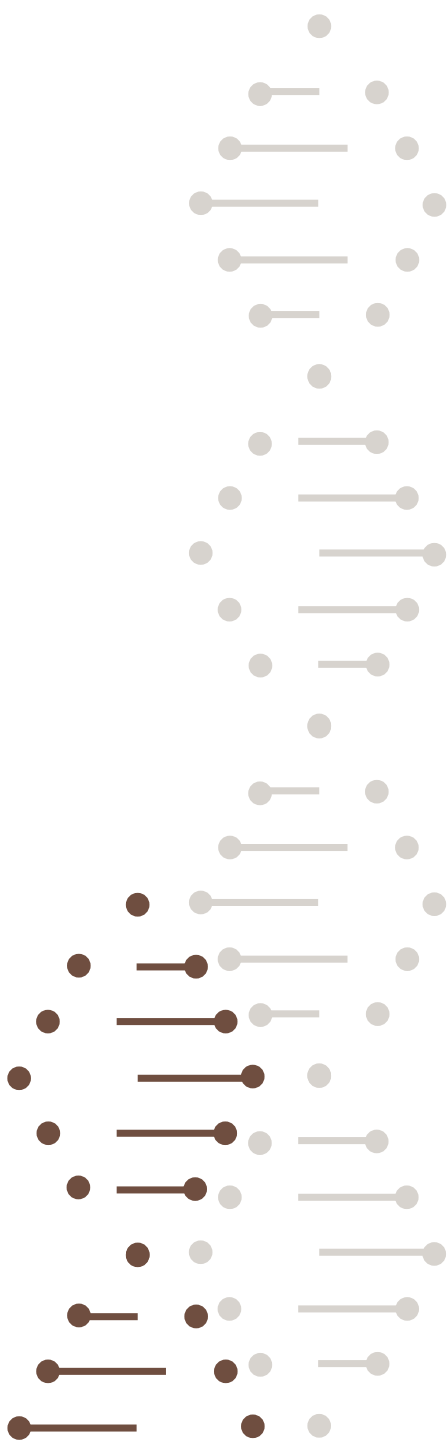
4.1. Cruzamiento de poblaciones en equilibrio de Hardy-Weinberg	101
4.2. Heterosis en cruzamientos interpoblacionales	105
4.2.1. Heterosis respecto al promedio de los progenitores	106
4.2.2. Heterosis respecto al mejor de los progenitores	107
4.3. Heterosis panmíctica y heterosis de línea base	108
4.4. Heterosis y epistasis	112

4.5. Modelo de Gardner y Eberhart sobre cruzamientos interpoblacionales	113
4.6. Medias de cruzamientos de prueba o <i>test crosses</i>	114
4.7. Habilidad combinatoria	117
4.7.1. Modelo de cruzamientos de prueba de Melchinger	120
4.8. Estimación y predicción de la media esperada de híbridos, variedades sintéticas y compuestos	121
4.8.1. Predictores BLUP (<i>Best Linear Unbiased Predictors</i>) del valor genotípico	122
4.8.1.1. Predicción BLUP y GBLUP del comportamiento de híbridos simples	122
4.8.1.2. Media esperada de híbridos de tres y cuatro líneas	125
4.8.2. Media esperada de variedades sintéticas y de compuestos	127
4.8.2.1. Variedades sintéticas	127
4.8.2.2. Compuestos	128
4.8.3. Modelo general de Vencovsky y Barriga (1992) para predecir medias de híbridos y de compuestos derivados de cruzamientos interpoblacionales	128

CAPÍTULO 5..... 133

Variabilidad poblacional

5.1. Variancia fenotípica y sus componentes	133
5.1.1. Abordajes en el estudio de la variabilidad fenotípica	134
5.2. Variabilidad genotípica	138
5.2.1. Cuantificación de la variancia genotípica de poblaciones experimentales derivadas del cruzamiento entre progenitores homocigotas	138
5.2.1.1. Variancia genotípica en poblaciones F_2	138
5.2.1.2. Variancia genotípica en poblaciones derivadas del retrocruzamiento de la F_1 a cada progenitor	139
5.2.1.3. Variancia genotípica en una población endocriada hasta homocigosis (de líneas recombinantes o haploides duplicados)	140
5.2.2. Variancia genotípica y sus componentes causales en una población no endocriada y bajo equilibrio de Hardy-Weinberg	140
5.2.2.1. Variancia aditiva y variancia de dominancia	141
5.2.2.2. Variancia epistática	143
5.3. La variancia genotípica y sus componentes causales en una población endocriada	144
5.4. Elección de probadores y variancia entre cruzamientos de prueba. Consideraciones teóricas e implicancias prácticas	149
5.5. Variancias genotípicas entre híbridos F_1	153
5.6. Variancia de cruzamientos interpoblacionales	153
5.7. Covariancia entre valores aditivos y de dominancia en poblaciones endocriadas y no endocriadas	155
5.8. Desequilibrio gamético	156
5.9. Heredabilidad	158
5.9.1. Error estándar de las estimaciones de heredabilidad	160



5.10. Utilidad de conocer los diferentes componentes de variancia genotípica en el mejoramiento	161
---	-----

CAPÍTULO 6..... 165

Medidas de parecido entre parientes

6.1. Causas del parecido entre parientes	165
Tipos de parientes	165
6.2. Cuantificación del parecido entre parecidos	167
6.2.1. Coeficiente de endogamia o de endocría. Identidad alélica por descendencia a nivel de individuos	168
6.2.2. Coeficiente de coancestría. Identidad alélica por descendencia a nivel de pares de individuos	169
6.2.2.1. <i>Coeficiente de coancestría en familias de medios hermanos</i>	171
6.2.2.2. <i>Coancestría entre hermanos completos</i>	172
6.2.2.3. <i>Coancestría entre progenie-progenitor</i>	173
6.2.2.4. <i>Coancestría entre tío y sobrino</i>	174
6.2.2.5. <i>Reglas para calcular el coeficiente de coancestría</i>	175
6.2.3. Coeficiente de Kempthorne	179
6.2.3.1. <i>Caso de hermanos completos</i>	179
6.2.3.2. <i>Caso de medios hermanos</i>	179
6.2.3.3. <i>Caso de progenie-progenitor</i>	180
6.3. Coeficiente de dominancia	181
6.4. Cálculo de coeficientes de parentesco en esquemas de pedigrí complejos mediante el método tabular	182

CAPÍTULO 7..... 187

Covariancia entre parientes

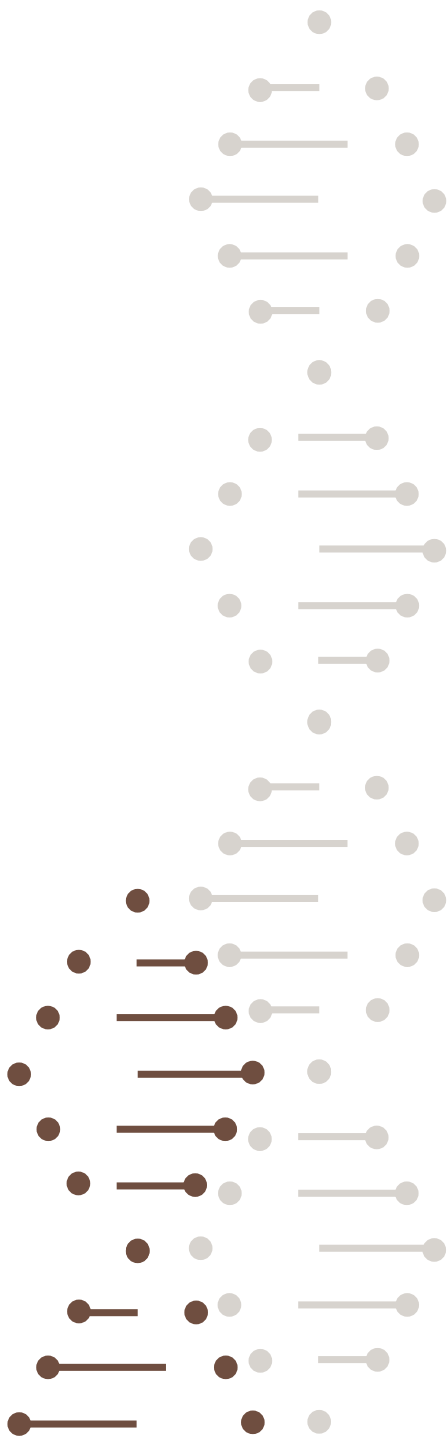
7.1. Covariancia genotípica entre individuos y sus componentes	187
7.1.1. Covariancia aditiva	190
7.1.2. Covariancia de dominancia	190
7.1.3. Covariancia genotípica de medios hermanos	192
7.1.4. Covariancia genotípica en sistemas de endocría aplicados a una población de referencia no endocriada	193
7.1.5. Modelo general de cómputo de la covariancia genotípica entre parientes pertenecientes a una población de referencia no endogámica	195
7.1.6. Modelo general de cómputo de la covariancia genotípica entre parientes pertenecientes a una población de referencia endogámica	196

CAPÍTULO 8..... 199

Diseños genéticos

8.1. Concepto y aplicación de los diseños genéticos	199
8.1.1. Modelos de efectos fijos y aleatorios	200
8.2. Consideraciones prácticas	202

8.3. Evaluación de genotipos en experimentos con repeticiones	202
8.4. Diseño de progenies biparentales o de hermanos completos	206
8.5. Diseño 1 de Carolina del Norte (NC1)	209
8.5.1. Análisis de NC1 en un ambiente de evaluación	211
8.5.2. Análisis de NC1 en varios ambientes de evaluación	213
8.5.3. Consideraciones sobre el diseño NC1	216
8.5.4. Casos especiales del NC1 en función de los objetivos y de las características de la población de referencia	217
8.5.4.1. Extensión del esquema del NC1 al desarrollo y evaluación de líneas endocriadas	217
8.5.4.2. Aplicación del NC1 en población de referencia derivadas del cruzamiento entre dos progenitores homocigotas	218
8.5.4.3. Aplicación del NC1 en poblaciones de referencia endocriadas	218
8.5.4.4. Consideraciones sobre epistasis en el diseño NC1	218
8.6. Diseño 2 de Carolina del Norte (NC2)	219
8.6.1. Análisis NC2 en un ambiente de evaluación	220
8.6.2. Análisis NC2 en varios ambientes de evaluación	223
8.6.3. Consideraciones sobre el diseño NC2	226
8.6.4. Casos especiales de aplicación de NC2 en función de los objetivos y de las características de la población de referencia	227
8.6.4.1. Extensión del esquema del NC2 al desarrollo y evaluación de líneas endocriadas	227
8.6.4.2. Aplicación del NC2 en poblaciones de referencia derivadas del cruzamiento entre dos progenitores homocigotas	228
8.6.4.3. Aplicación del NC2 en poblaciones de referencia endocriadas	228
8.6.4.4. Consideraciones sobre epistasis en el diseño NC2	229
8.7. Diseño 3 de Carolina del Norte (NC3)	229
8.8. Diseños dialélicos	233
8.8.1. Diseño dialélico de Griffing (1956)	235
8.8.1.1. Método 1	241
8.8.1.2. Método 2	242
8.8.1.3. Método 3	244
8.8.1.4. Método 4	246
8.8.2. Variancia de los efectos	247
8.8.4. Importancia relativa de las habilidades combinatorias general y específica	248
8.9. Diseños dialélicos de Gardner y Eberhart	249
8.9.1. Análisis I	249
8.9.2. Análisis II	250
8.9.3. Análisis III	254
8.9.4. Relación entre los análisis II y III	254
8.10. Procedimientos para estimar los componentes de variancias poblacionales	255
8.10.1. Función de máxima verosimilitud	256
8.10.1.1. Algunas limitaciones del análisis de máxima verosimilitud	256





CAPÍTULO 9.....259 **Selección**

9.1. Tipos de selección	260
9.2. Selección artificial	261
9.3. Algunas definiciones en relación con la selección artificial	262
9.4. Respuesta a la selección direccional	263
9.4.1. Diferencial de selección e intensidad de selección	265
9.4.1.1. Apartamiento de la distribución poblacional respecto de la distribución normal	270
9.5. Cambios de frecuencia génica por ciclo de selección	272
9.6. Respuesta a diferentes métodos de selección	276
9.7. Factores que afectan la continuidad de la respuesta a través de ciclos de selección	281
9.8. Selección entre y dentro de familias	282
9.9. Repetibilidad de la respuesta a la selección Heredabilidad realizada	284
9.10. Respuesta asimétrica a la selección	285
9.11. Límite a la selección	287
9.12. Número de factores efectivos y respuesta a la selección	294
9.13. Estimación de respuesta a la selección	294
9.14. Identificación y selección de progenitores de poblaciones de cría	295
9.14. 1. Modelos de Dudley	295

CAPÍTULO 10.....307 **Selección recurrente**

10.1. Objetivo de la selección recurrente	307
10.2. Clasificación de los métodos de selección recurrente	312
10.3. Predicción de respuesta a diferentes métodos de selección recurrente	314
10.3.1. Procedimiento de predicción de respuesta de Eberhart	315
10.3.2. Procedimiento de predicción de respuesta de Vencovsky	316
10.4. Respuesta a la selección recurrente intrapoblacional	317
10.4.1. Selección recurrente fenotípica	317
10.4.2. Selección recurrente espiga x hilera (procedimiento original)	319
10.4.3. Selección recurrente espiga x hilera modificada	321
10.4.4. Selección recurrente espiga x hilera doble modificada	322
10.4.5. Otros métodos de selección de medios hermanos Selección recurrente entre cruzamientos de prueba (Selección recurrente por habilidad combinatoria general y selección recurrente por habilidad combinatoria específica)	323
10.4.6. Selección recurrente de hermanos completos	326
10.4.7. Selección recurrente de familias endocriadas $S_{0,1}$	327
10.4.8. Selección recurrente de familias endocriadas $S_{1,2}$	328
10.5. Respuesta a la selección recurrente recíproca (interpoblacional)	329

10.5.1. Selección recurrente de medios hermanos interpoblacionales o selección recurrente recíproca de medios hermanos	330
10.5.2. Selección recurrente recíproca de hermanos completos	334
10.6. Selección recurrente, número efectivo y endogamia	337
10.7. Selección recurrente genómica	340
10.7.1. Características de la población de entrenamiento	345
10.7.2. Complejidad del carácter bajo selección y precisión del modelo	346
10.7.3. Efecto del tamaño efectivo sobre la precisión del modelo	346
10.7.4. Número y tipo de marcadores necesarios y precisión del modelo	347
10.7.5. Densidad de marcadores y precisión del modelo	348
10.7.6. Heredabilidad del carácter y precisión	348
10.7.7. Desequilibrio de ligamiento y precisión del modelo	348
10.7.8. Respuesta a la selección genómica	348

CAPÍTULO 11 351

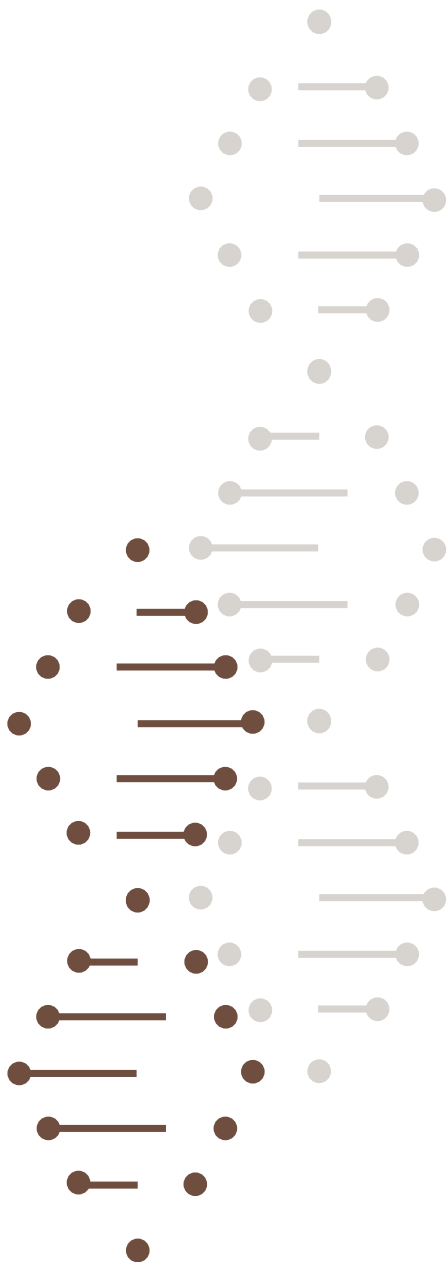
Correlación entre caracteres

11.1. Estimación de la correlación fenotípica	356
11.2. Estimación de la correlación genética	357
11.2.1. Estimación a partir de evaluación de familias	358
11.2.2. Estimación a partir de diseño padre-progenie	358
11.2.3. Estimación a partir de experimentos de selección (o respuesta correlacionada)	359
11.3. Precisión de las estimaciones de correlaciones y relación con el tamaño muestral	362
11.4. Correlación genética entre resultados de evaluaciones de genotipos en etapas tempranas y tardías	364

CAPÍTULO 12 369

Selección de múltiples caracteres

12.1. Selección en tándem	369
12.2. Selección por niveles de cortes independientes	371
12.3. Índices de selección	373
12.3.1. Ecuaciones del índice de selección	376
12.3.2. Propiedades del índice y del genotipo agregado	377
12.3.3. Respuesta a la selección en el genotipo agregado	379
12.3.4. Expresión de los índices en términos matriciales	382
12.3.5. Pesos del índice	385
12.3.6. Respuesta a la selección basada en índices de selección	385
12.3.7. Índice óptimo e Índice estimado	386
12.3.7.1. Empleo de caracteres secundarios o variables indicadoras	388
12.3.7.2. Índice óptimo de Henderson	390
12.3.8. Índice base o de Williams	392
12.3.8.1. Índice base modificado	392
12.3.9. Índices restringidos	393



12.3.9.1. Índices restringidos de Kempthorne y Nordskog (1959), Cunningham et al. (1970) y Tallis (1962)	394
12.3.9.2. Índice de ganancias o respuestas deseadas (Pesek y Baker, 1969; Harville, 1975; Tai, 1977)	395
12.3.10. Índice multiplicativo	396
12.3.11. Índice de suma de rankings	397
12.3.12. Otros índices no lineales	397
12.4. Comparación de la eficiencia de diferentes estrategias de selección para caracteres múltiples	398
12.5. Selección combinada basada en información molecular y fenotípica	400
12.5.1. Índice Marcador-Fenotipo	401
12.6. Aplicación del concepto de índices de selección a la selección intrapoblacional en varias etapas	402
12.6.1. Respuestas con diferentes estrategias de selección por etapas	405
12.6.1.2. Selección en una etapa	405
12.6.1.3. Selección en dos etapas	407
12.7. Selección con base en información de parientes	410
12.8. Métodos para la elección de progenitores basados en caracteres múltiples	412

CAPÍTULO 13.....415

Interacción genotipo x ambiente

13.1. Introducción	415
13.2. Modelo con interacción genotipo x ambiente	416
13.3. Interacción genotipo x ambiente y predicción de la respuesta a la selección	422
13.4. Experimentos multiambientales	423
13.5. Modelos estadísticos de análisis de la interacción genotipo x ambiente	426
13.6. Adaptabilidad, estabilidad y confiabilidad genotípicas	432

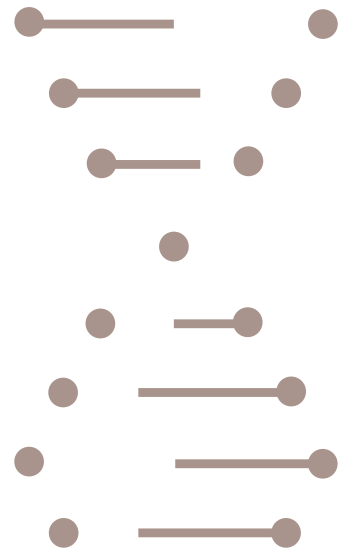
CAPÍTULO 14.....435

Ploidía y herencia cuantitativa

14.1. Introducción	435
14.2. Variabilidad genotípica	437
14.3. Poliploidía y equilibrio poblacional de Hardy-Weinberg	440
14.4. Modelo genético	442
14.5. Variancias y covariancias en especies autoploiploides	444
14.6. Poliploidía y endogamia	446

BIBLIOGRAFÍA.....449

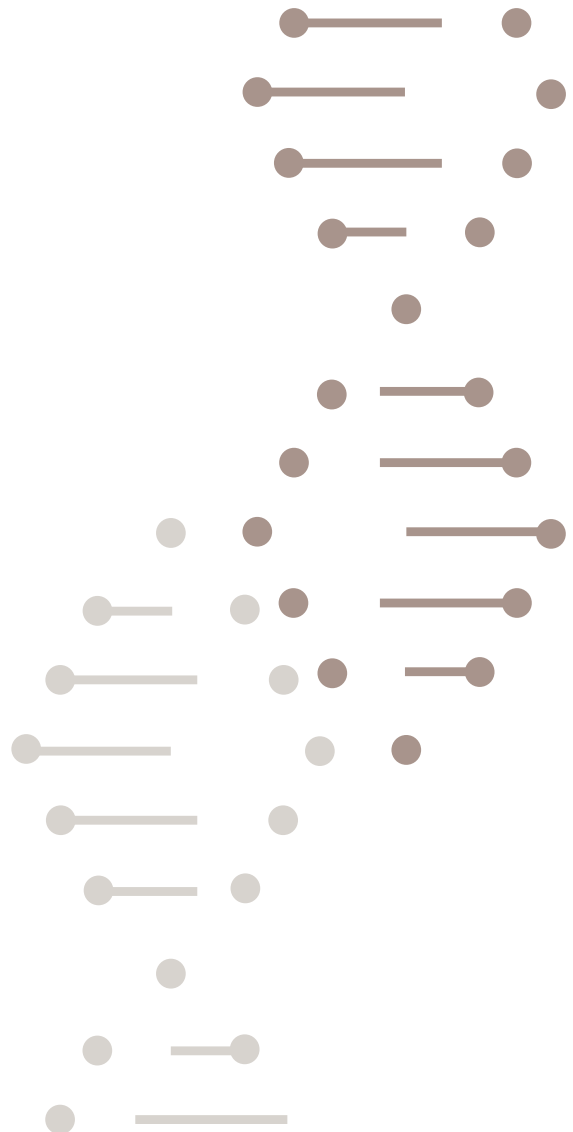
ANEXO	461
Anexo - Índice	482

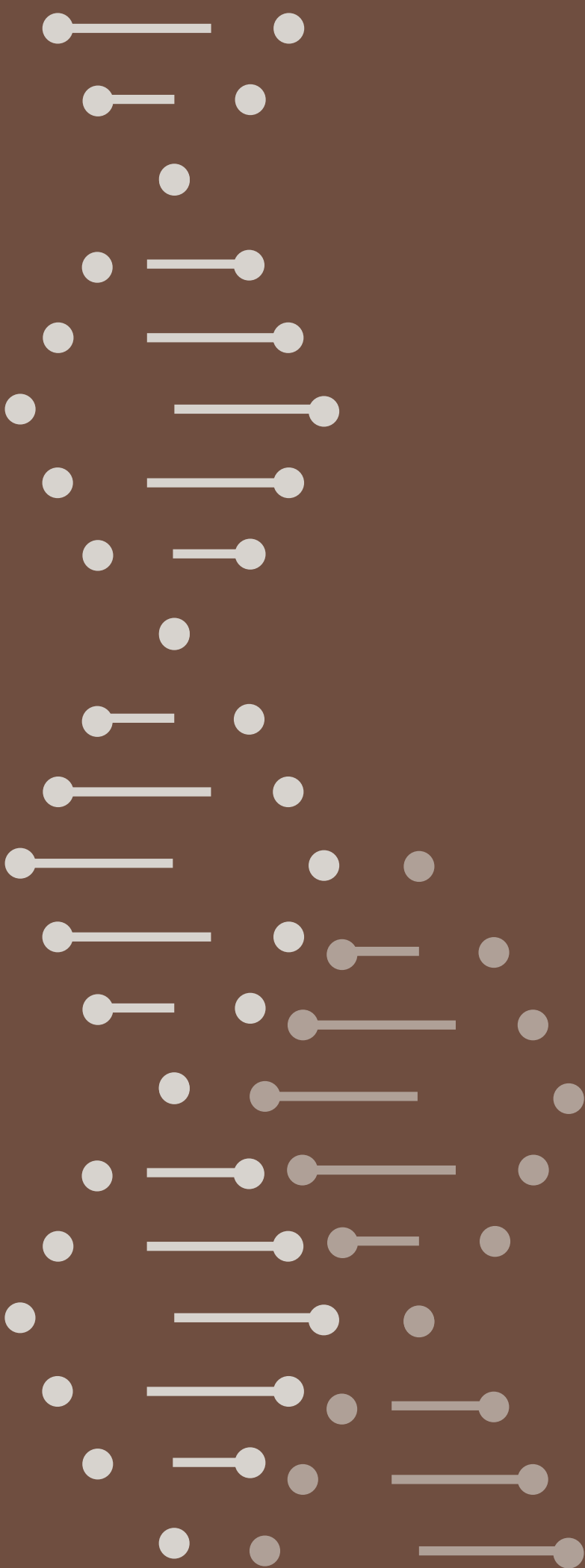


A mis hijos Andrés, Ana y Alicia, a mis nietos Agustín,

Benjamín, Ignacio y Delfina, a Mabel.

A la memoria de mis padres, Carmen y Oscar.





PRÓLOGO



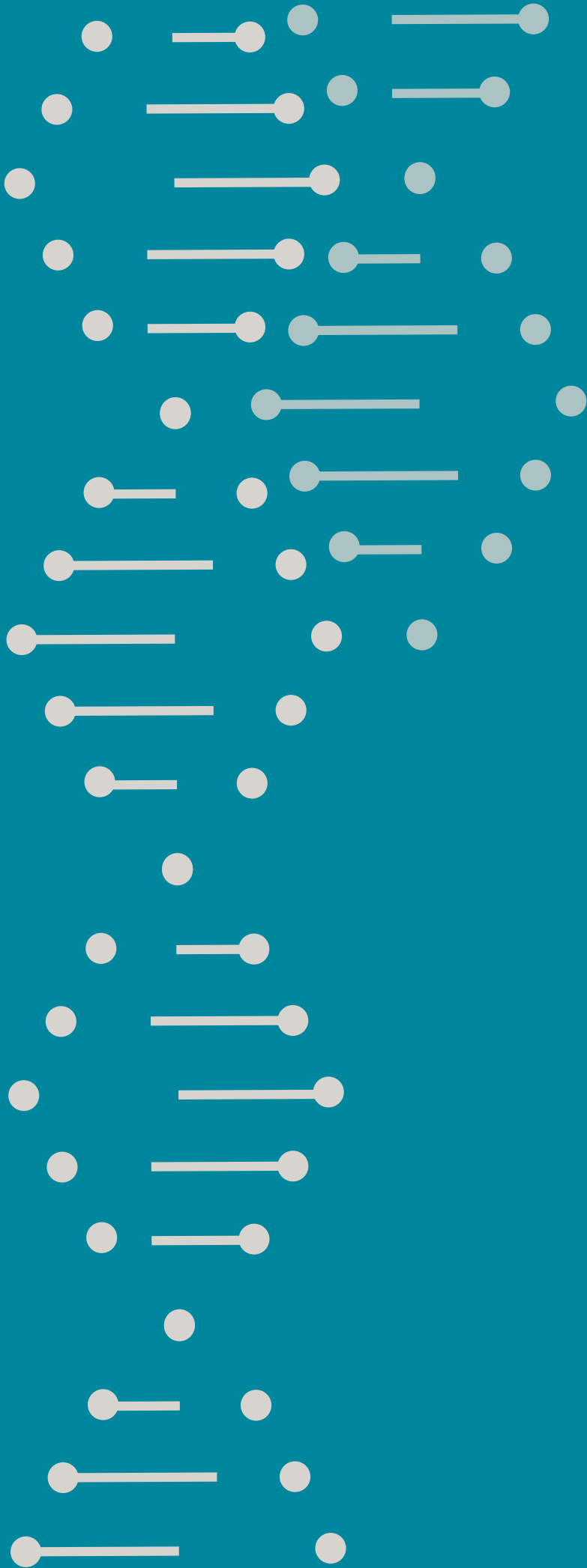
Este texto está dirigido a estudiantes de grado y posgrado en Agronomía, Biología, Genética, y Genética Aplicada, con el objetivo de acercar los conceptos fundamentales de la Genética Cuantitativa que dan sustento científico a los métodos de selección, tanto fenotípica como combinada con información genotípica. Intenta constituirse también en una guía de consulta para la práctica profesional de los fitotecnistas.

La secuencia de los contenidos que propone cada uno de los catorce capítulos guarda relación con la organización lógica de un curso de posgrado en Genética Cuantitativa General. Para quienes posean una formación previa en Genética de Poblaciones, la lectura de algunos contenidos puede obviarse, como las secciones dedicadas al equilibrio de Hardy-Weinberg, los conceptos de tamaño efectivo, deriva genética y endocría, o el de parecido entre parientes.

Este libro es resultado, por supuesto perfectible, de mi formación y experiencia profesional en el fitomejoramiento dentro del Instituto Nacional de Tecnología Agropecuaria, y más recientemente en la docencia de grado y posgrado en la Universidad Nacional del Noroeste de Buenos Aires y la Universidad Nacional de Rosario. A él han contribuido, en buena medida, los intercambios de experiencias con mis profesores y colegas talentosos con los que tuve la fortuna de estudiar o trabajar, así como de alumnos a quienes orienté o dirigí en su paso por la universidad pública. A todos ellos mi agradecimiento. Finalmente un reconocimiento especial al Dr. Luis Narro León, al Dr. César Gabriel López, a la Mg.Sc. Erika Mroginski y a la Dra. Teresa Boca por su revisión del manuscrito, sugerencias y comentarios sobre el texto.

Ing. Agr. Guillermo Hugo Eyhérbide, M.Sc., Ph.D.

Pergamino, 12 de febrero de 2019



Genética cuantitativa y variación continua

1.1. La variación continua

El estudio de las características o caracteres genéticos requiere el registro de su expresión a nivel de individuos, familias o poblaciones. Con frecuencia es posible agrupar a los individuos según la expresión visual o numérica de tales caracteres en unas pocas clases fácilmente distinguibles o reconocibles. Es el caso, por ejemplo, de los caracteres estudiados por Mendel (1865). A diferencia de la segregación en clases distintivas de ciertos caracteres, existen otros (la inmensa mayoría, y especialmente los de mayor importancia económica) cuya expresión no se limita a unas pocas categorías, sino que su variación en la expresión no presenta discontinuidades marcadas. Esta última clase de variación se la califica como continua y los caracteres que la presentan se denominan caracteres cuantitativos, métricos o poligénicos.

1.2. La Genética Cuantitativa


La rama de la Genética que estudia los caracteres métricos es la Genética Cuantitativa. Existen otras denominaciones como Genética Estadística o Genética Biométrica que pueden considerarse sinónimos. Más recientemente con el desarrollo de la Genómica y el empleo de marcadores moleculares se generaron conceptos de lo que se denomina Estadística Genómica, que se enfoca en la modelación y en el tratamiento de datos moleculares.

El objetivo de la Genética Cuantitativa es proveer modelos mecánicos que ayuden a entender el proceso de evolución. Para ello es necesario comprender la esencia de la variación continua. Esta disciplina posee aplicaciones muy importantes en estudios de evolución ya que permite dilucidar la arquitectura de los caracteres cuantitativos, inferir sus causas y especular sobre su probable evolución. Debido a que muchos de los caracteres cuantitativos poseen importancia económica, la Genética Cuantitativa se aplica en el mejoramiento genético de esas características, mediante la selección artificial de animales y vegetales. En este campo de aplicación, la Genética Cuantitativa permite predecir la respuesta a la selección y diseñar las estrategias selectivas que maximicen esa respuesta. En los campos de la Genética y Genómica Humanas, la Genética Cuantitativa permite asociar características genotípicas con condiciones de riesgo para contraer enfermedades y contribuye a adoptar medidas preventivas con impacto positivo sobre la salud.

Dos fenómenos genéticos referidos a las propiedades de los caracteres cuantitativos y que son muy importantes en estudios de herencia y de selección artificial son el parecido entre parientes y la depresión por endocria y la heterosis, tópicos que recibirán atención en las secciones respectivas.

1.3. Reseña histórica de la Genética Cuantitativa

Los primeros trabajos sistemáticos de selección de plantas datan de los siglos XVIII y XIX, fundamentalmente los llevados a cabo por Van Mons en Bélgica, Knight (1799) en Inglaterra y Cooper en Estados Unidos de América. Las contribuciones fundacionales de la Genética Cuantitativa abarcaron un período extenso, desde 1850 hasta 1930, aproximadamente. De Vilmorin (1850) acuñó el término “principio de aislamiento” y el primer antecedente de la utilización de pruebas de progenie para estimar y comparar el mérito genético de los individuos de una población. Su método suponía la selección de líneas con base en el comportamiento de sus descendientes. De Vilmorin también investigó sobre las diferencias de respuestas a la selección que presentaban las plantas alógamas respecto a las autógamias, lo cual constituyó un antecedente relevante para comprender las futuras experiencias de Johansen (1903).



Darwin (1889) intentó explicar por qué el fenotipo de la progenie era intermedio al de sus progenitores, circunstancia que atribuyó a la “herencia de la mezcla” (*blending inheritance*). Este principio se contraponía con la observación de que la variabilidad no se agotaba con el transcurso de las generaciones. Recién medio siglo más tarde, con los trabajos de Yule (1906), se descubre la segregación independiente de factores múltiples y su papel como mecanismo de mantenimiento de la variabilidad. Darwin consideraba a la evolución como la consecuencia de la acumulación de pequeños cambios en caracteres adaptativos producidos por la selección natural.

Galton (1889) publicó sus estudios sobre la relación entre la estatura de los progenitores y de su progenie en poblaciones humanas. Fue el creador del término regresión. Galton experimentó un relativo fracaso por no llegar a entender las relaciones genéticas progenitor-progenie, debido a la naturaleza cuantitativa del carácter y al empleo de familias de tamaño reducido en sus experimentos. Se esforzó en clasificar las formas variables en grupos a la manera mendeliana y describirlos en términos estadísticos. La regresión padre-progenie encontrada no respetaba la relación 1:1 entre la estatura promedio de los progenitores y la estatura promedio de sus hijos, sino que la pendiente de la recta de regresión era inferior a 1. Galton llamó a este fenómeno “regresión hacia la mediocridad” y creyó que actuaba en contra de las posibilidades de evolución por selección. Sostuvo que para que la evolución tuviera lugar serían necesarias mutaciones de efecto grande, más que la selección actuando sobre la variación continua. Karl Pearson (1904) advirtió el error en la lógica de Galton (la selección actúa en cada generación sobre una media poblacional nueva) y realizó aportes fundamentales al tratamiento estadístico de datos fenotípicos utilizando regresión y correlación y se lo reconoce como uno de los mayores exponentes de la escuela biométrica.

Hacia 1900 se produjo el redescubrimiento de las leyes de Mendel y se renovó el interés en estudiar la herencia de caracteres cuya expresión fenotípica era discreta y fácilmente discernible en clases o grupos bien definidos. Bateson (1902) es uno de los referentes de la escuela mendeliana.


El debate entre las escuelas biométrica y mendeliana se daba respecto a si los caracteres discretos y los caracteres continuos poseían la misma herencia y relación con la evolución. Para la escuela mendeliana, el surgimiento de nueva variabilidad se debía a mutaciones de gran efecto, en tanto que para la escuela biométrica la evolución resultaba de la selección natural actuando sobre caracteres continuos. Los métodos empleados por Mendel y por Galton no pudieron por sí solos explicar la herencia de ambos tipos de caracteres. La síntesis y el fin del debate ocurrieron años después, producto de las experiencias de Johannsen (1903), Nilsson-Elhe (1909), Emerson (1910) y East (1910), entre otros.

En 1903 Johannsen publicó su trabajo, reconocido luego como *teoría de las líneas puras*. Partiendo de una muestra de una población de semilla de poroto

(*Phaseolus vulgaris*) que presentaba heterogeneidad para el tamaño (peso) de la semilla, seleccionó semillas de diferentes tamaños y las multiplicó separadamente. Observó que la descendencia de semillas grandes tendía a ser más pesada que la descendencia de las semillas pequeñas, es decir, que la selección había sido efectiva. Pudo así establecer 19 linajes o líneas con un tamaño de semilla característico a partir de las 19 semillas de la muestra original inicialmente seleccionadas por tamaño. Johannsen concluyó que el lote original de semillas de poroto era una mezcla mecánica de "líneas puras", cada una con un peso característico. Dentro de cada una de las 19 líneas había cierta variabilidad en el peso de la semilla, pero esta variabilidad era menor que la variabilidad que existía entre las 19 líneas. Johannsen postuló que la variabilidad o variación entre las líneas era de origen genético y que la variación dentro de las líneas se debía a causas no genéticas. Para probarlo realizó nuevos experimentos de selección por tamaño de semilla, pero esta vez dentro de cada una de las líneas puras y observó que la selección no era efectiva. *Phaseolus vulgaris* es una especie autó-gama, por lo cual la homocigosis previa generó distintas líneas puras presentes en la muestra original y que pudieron seleccionarse. En cambio, la variación total observable dentro de cada una de esas líneas era producto del ambiente. Se concluyó que la variabilidad observada en el tamaño de semilla respondía tanto a causas heredables como no heredables y que la manera de conocer su contribución relativa solo podría determinarse mediante pruebas de progenies. La experiencia de Johannsen contribuyó a comprender la relación entre genotipo y fenotipo: la discontinuidad en la distribución de los genotipos podía suavizarse y convertirse en variación fenotípica continua si a la distribución genotípica se le incorporaba la distribución de los efectos ambientales.

Nilson-Ehle (1909) analizó la segregación de filiales resultantes del cruzamiento de líneas de trigo y de avena que diferían en el color del grano y concluyó que varios genes segregando en forma independiente y simultánea podían producir la variación continua del color. Encontró gradaciones en el color de grano de trigo, desde blanco a rojo, producto de varios genes cuyas segregaciones individuales podían analizarse separadamente como un carácter mendeliano. Según la hipótesis de los factores múltiples, un número suficiente de genes con efectos individuales pequeños, similares y segregando conjuntamente podían causar variación continua. Cada uno (factor o gen) se heredaba de manera mendeliana. El efecto de los factores sobre el carácter era acumulativo (o de dosis) y los tipos intermedios eran los más frecuentes. La distribución del carácter se asemejaba a la curva normal en la medida que al efecto de las segregaciones se le sumara el efecto ambiental. Otras contribuciones a la hipótesis de los factores múltiples se debieron a Shull (1908) y Emerson (1910). East (1910), por su parte, comprobó la hipótesis de factores múltiples en maíz.

Ronald Fisher (1918) y Sewald Wright (1921) integraron conocimientos de la Genética y la Estadística para entender la herencia de los caracteres cuantitativos. La variación de tipo continuo se explicaba por la segregación simultánea de mu-



chos genes, cada uno de pequeño efecto y siguiendo las leyes mendelianas (modelo infinitesimal). Utilizaron los datos de la escuela biométrica para demostrar la dominancia en factores múltiples e intentaron la primera partición de la variación continua en componentes heredables y no heredables (o ambientales). Los principios básicos de la Genética Cuantitativa fueron delineados por Fisher y Wright, que sentaron las bases teóricas para el mejoramiento animal y vegetal y también ayudaron a entender la herencia de desórdenes genéticos en humanos, tales como psicopatologías, enfermedades cardíacas y malformaciones congénitas. Los aportes a la Genética Cuantitativa trascendieron a otras ramas de la ciencia, como la Estadística. A manera de ejemplos, el desarrollo de la teoría de la regresión y correlación de Pearson se inspiró en los trabajos de Galton, y Fisher introdujo el concepto de partición de la variancia total en componentes, sobre los cuales se basan las técnicas de análisis de variancia, a su vez un hito en el diseño experimental.

1.4. Tipos de caracteres

1.4.1. Caracteres cualitativos, mendelianos o discretos

Su variación se refleja en clases distintivas o discretas y su estudio se basa en recuentos y proporciones. Están controlados por uno o pocos genes y como el ambiente ejerce escaso efecto sobre su expresión fenotípica, existe una estrecha relación entre genotipo y fenotipo. Ejemplos de este tipo de caracteres son el color de flor, el color de hilo de la semilla, la resistencia a ciertas enfermedades, etc.

1.4.2. Caracteres cuantitativos, métricos, continuos o poligénicos

Presentan variaciones de grado, sin discontinuidades. Requieren una escala de medición y para describirlos se emplean parámetros poblacionales o estimadores muestrales (medias, variancias, etc.). Su expresión está sujeta a considerable efecto ambiental. Ejemplos de este tipo de caracteres son la inteligencia, la estatura, el rendimiento de grano, la ganancia de peso, la producción de leche, etc. Gran parte de los caracteres de importancia económica son de tipo cuantitativo.

Debido a que la segregación ocurre en un número grande de loci con efectos pequeños y similares (poligenes), los efectos de sustitución de alelos en diferentes loci resultan triviales en relación con la cantidad de variabilidad observable (Lerner, 1958). Por lo tanto los efectos fenotípicos de sustitución alélica son intercambiables. Dicho de otro modo, idéntica expresión

fenotípica puede resultar de distintos genotipos, aún en ausencia de dominancia, epistasia y efectos ambientales. De manera similar, Comstock (1978) clasificó un carácter como cualitativo cuando el efecto de sustituir un alelo por otro es relativamente grande en relación con la variación fenotípica y como cuantitativo cuando el efecto de sustitución es relativamente pequeño en relación con la variación fenotípica total. Para que ocurra esta situación debe haber muchos genes involucrados en el control del carácter o bien el efecto del ambiente debe ser de una magnitud importante como para enmascarar el efecto genotípico (Figura 1.1).

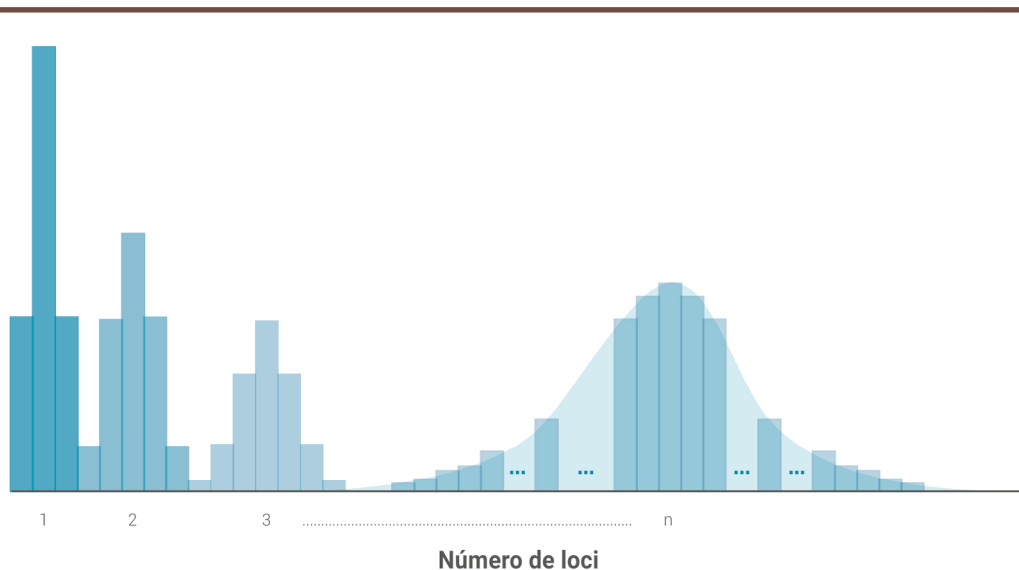



Figura 1.1. Ejemplo de distribución de clases fenotípicas en función del número de loci segregantes (n) para un carácter en una filial F_2 . Se asume que los alelos tienen efecto aditivo y equivalente entre loci. El efecto ambiental contribuye a la aproximación a la distribución normal cuando n es grande.

Los poligenes pueden tener efecto pleiotrópico (sobre dos o más caracteres) y actuar como modificadores o supresores de otros genes. La falta de correspondencia entre fenotipo y genotipo se debe no solo por el menor efecto de sustitución de alelos, sino también por la mayor importancia del efecto del ambiente sobre la expresión. Idéntico fenotipo puede resultar de distintos genotipos, o diferentes fenotipos pueden responder a un mismo genotipo. A manera de ejemplo, comparemos el rendimiento de un cultivar híbrido simple susceptible a una enfermedad evaluado en un ambiente donde esa enfermedad está presente con el del mismo híbrido cultivado en otro ambiente sin enfermedad. En este caso tendremos dos fenotipos diferentes (rendimientos diferentes) para el mismo genotipo (susceptible a la enfermedad). Consideremos ahora el comportamiento de dos híbridos de similar potencial de rendimiento en condiciones no limitadas por enfermedades, uno genéticamente resistente y otro susceptible a una enfermedad,



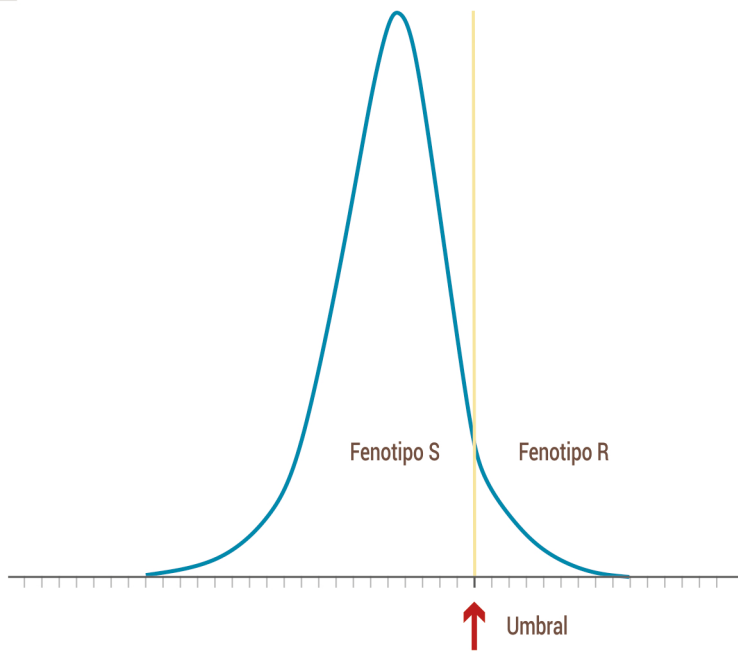
cuando se evalúan conjuntamente en un ambiente libre de la enfermedad. Estaremos en presencia de un mismo fenotipo (similar rendimiento) para dos clases de genotipos diferentes (susceptibles y resistentes).

Debido a que la segregación de los genes que contribuyen a un carácter métrico no puede seguirse individualmente, se necesitan técnicas y métodos específicos para su estudio. Los caracteres cuantitativos pueden tomar cualquier valor entre dos extremos de una escala de medición, pero el número de fenotipos detectables puede estar limitado por la capacidad o precisión con la que pueden medirse. Si idealmente tuviéramos un instrumento de medición perfecto, tal vez sería posible detectar todas las discontinuidades de clases, conforme los principios de la genética mendeliana.

Algunos caracteres tienen la particularidad de medirse en números enteros, pero se los clasifica como cuantitativos porque también están controlados por muchos genes de efectos pequeños e influenciados por el ambiente y se los denomina merísticos. Ejemplo de ellos son el tamaño de camada, el número de granos por espiga, el número de espigas por planta en maíz, etc.

Los caracteres umbrales son también cuantitativos, pero su expresión toma valores de tipo cualitativo, como presencia/ausencia, sano/enfermo, capacitado/discapitado, etc. En realidad dependen de la expresión de una característica subyacente (riesgo, fertilidad, estabilidad, etc.) que se asume con distribución continua. A partir de determinado valor del carácter subyacente (umbral), el fenotipo recibe una clasificación discreta (ej. sano/enfermo). El número de clases fenotípicas depende del número de valores umbrales. Así, con un umbral pueden distinguirse dos clases fenotípicas y con dos valores umbrales, tres clases fenotípicas (Figura 1.2).

[a]



[b]

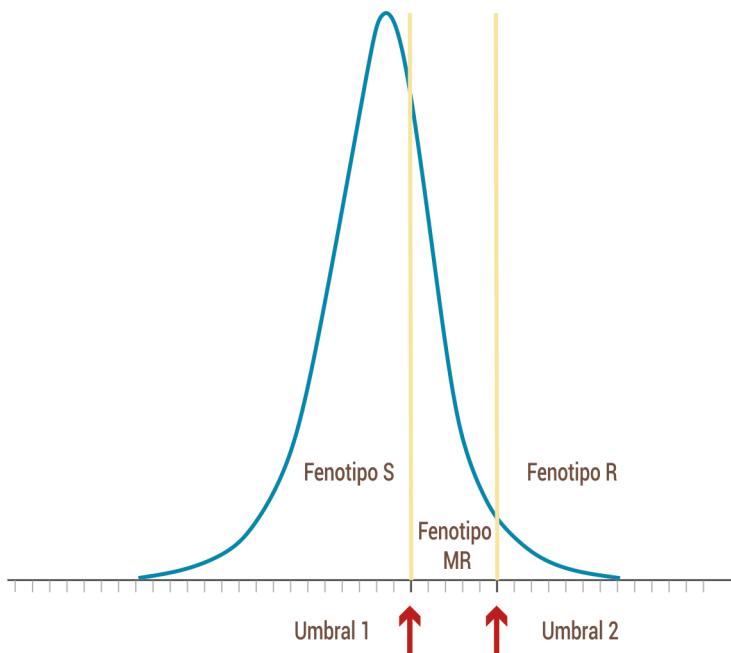
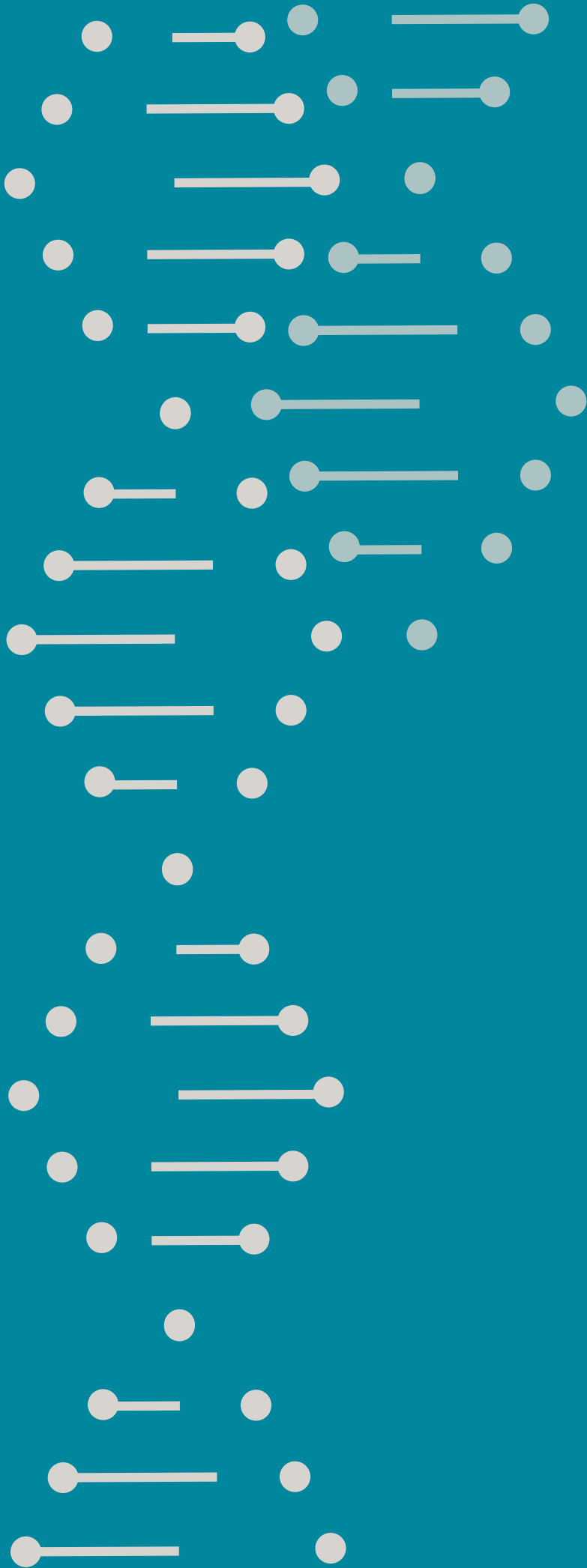


Figura 1.2. Distribución de valores fenotípicos para un carácter poligénico subyacente. [a] con un umbral y dos clases fenotípicas [ejemplo: susceptible (S) y resistente (R)] y [b] con dos umbrales y tres clases fenotípicas [ejemplo: susceptible (S), moderadamente resistente (MR) y resistente (R)].



1.5. Arquitectura de los caracteres cuantitativos

La arquitectura de un carácter hace referencia al número de genes que lo afectan, su ubicación, la magnitud y la distribución de sus efectos génicos y la importancia relativa de los diferentes modos de acción génica (Holland, 2007). La arquitectura de los caracteres cuantitativos debe ser muy tenida en cuenta en la decisión del método de mejoramiento para aplicar, especialmente en el caso de la selección genómica.



Genética de poblaciones

2.1. Población genética

El término población se refiere a la totalidad de un conjunto o grupo de individuos. En el caso de poblaciones genéticas nos referimos al conjunto o grupo completo de individuos que comparten un mismo acervo o pool génico, por cuanto se reproducen entre sí, comparten un nicho ecológico en un espacio geográfico y tiempo determinados y que logran continuidad de generación en generación.

En la naturaleza tanto las poblaciones vegetales como los animales son de tamaño finito, pero ese tamaño puede ser lo suficientemente elevado como para que resulte impracticable estudiarlas a partir de la totalidad de los individuos que la componen. Por ello la descripción, la caracterización y la evaluación de las poblaciones se realizan sobre muestras poblacionales.

Estas son subconjuntos de individuos de la población a partir de las cuales se hacen inferencias sobre las propiedades de la población completa. Para hacerlo correctamente, las muestras deben ser representativas de la población. Para asegurar tal representatividad los individuos que la componen deben estar en número suficiente para dar cuenta de la variabilidad de la población y debieron haber sido escogidos aleatoriamente.

Las poblaciones se describen empleando parámetros estadísticos de posición y de dispersión (ver Anexo), fundamentalmente, la media y variancia paramétricas para uno o más caracteres. En este último caso también resulta pertinente determinar la existencia y magnitud de la asociación entre los caracteres.

2.2 Poblaciones genéticas y equilibrio de Hardy-Weinberg

Para cada carácter, las poblaciones genéticas tienen frecuencias génicas o alélicas en cada loci que les resultan características en el tiempo, siempre que a través de las sucesivas generaciones no operen procesos que puedan afectarlas. Esos procesos se denominan sistemáticos, como la migración, mutación y selección y procesos dispersivos aleatorios como la deriva genética. El lector puede obtener una descripción detallada de estos procesos y sus efectos sobre las frecuencias génicas en Falconer y Mackay (2001).

En ausencia de tales procesos sistemáticos y dispersivos, las frecuencias génicas en una población se mantienen sin cambio entre generaciones sucesivas y las frecuencias genotípicas, a su vez, quedan determinadas por las frecuencias génicas (Figura 2.1). Se dice entonces que esa población está en la condición de equilibrio descrita por la Ley de Hardy-Weinberg (EHW) (Hardy, 1908; Weinberg, 1908). Por una parte, cuanto menor sea la frecuencia del alelo favorable, mayor será la proporción de ellos que se encuentran en los genotipos heterocigotas. Por otra parte, la mayor proporción de individuos de la población serán heterocigotas cuando las frecuencias génicas sean $p=q=0,5$.

Asumiendo que en una población de una especie diploide las frecuencias génicas para un locus con dos alelos (A^+ , A^-) son p y q , respectivamente, las frecuencias genotípicas en la condición de EHW serán p^2 para A^+A^+ , q^2 para A^-A^- y $2pq$ para el heterocigota A^+A^- .

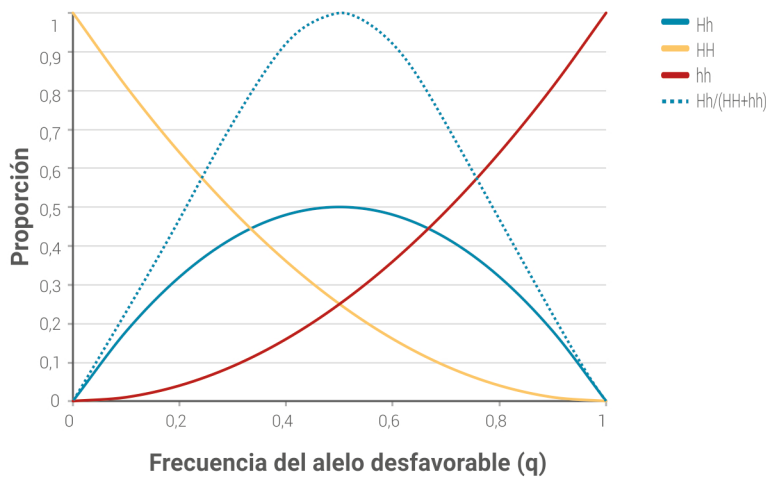


Figura 2.1. Proporción de individuos homocigotas para el alelo favorable (HH), para el desfavorable (hh), de individuos heterocigotas (Hh) y proporción de heterocigotas a homocigotas [Hh/(HH+hh)] en función de la frecuencia del alelo desfavorable (q).

Estas frecuencias genotípicas en el equilibrio corresponden al desarrollo del cuadrado del binomio de las frecuencias génicas $(p + q)^2$. En el caso de un locus con alelos múltiples, las frecuencias genotípicas en el EHW pueden calcularse a partir del cuadrado de un trinomio si existen tres alelos posibles, o del cuadrado de un polinomio de orden n, siendo n el número de alelos presentes en la población en dicho locus. Por ejemplo, en el caso de cuatro alelos por locus con frecuencias génicas p, q, r, s, las frecuencias genotípicas en el equilibrio serán:

$$\begin{aligned} (p + q + r + s)^2 &= (p + q + r + s)(p + q + r + s) = \\ &= p^2 + 2pq + 2pr + 2ps + q^2 + 2qr + 2qs + r^2 + 2rs + s^2 \end{aligned}$$

Una expresión general y equivalente para el caso de alelos múltiples es (Wricke y Weber, 1986):

$$\left(\sum_{i=1}^n p_i G_i \right)^2$$

en la que n es el número de alelos que varía de 1 a n, y p_i es la frecuencia del alelo G_i .

Por ejemplo cuando $n=5$:


$$\begin{aligned} \left(\sum_{i=1}^n p_i G_i \right)^2 &= p_1^2 G_1 G_1 + p_2^2 G_2 G_2 + p_3^2 G_3 G_3 + p_4^2 G_4 G_4 + p_5^2 G_5 G_5 + \\ &+ 2p_1 p_2 G_1 G_2 + 2p_1 p_3 G_1 G_3 + 2p_1 p_4 G_1 G_4 + 2p_1 p_5 G_1 G_5 + \\ &+ 2p_2 p_3 G_2 G_3 + 2p_2 p_4 G_2 G_4 + 2p_2 p_5 G_2 G_5 + 2p_3 p_4 G_3 G_4 + \\ &+ 2p_3 p_5 G_3 G_5 + 2p_4 p_5 G_4 G_5. \end{aligned}$$

Una población que no esté en la condición de EHW puede alcanzarlo siempre que exista apareamiento al azar entre los individuos y no operen los procesos sistemáticos ni dispersivos.

La mayoría de las poblaciones naturales de especies de polinización cruzada o alógamas se encuentran en EHW y sus poblaciones experimentales lo están o podrán estarlo en una generación cuando consideramos un simple locus. Cuando hay más de un locus que controla el carácter, el desequilibrio gamético puede demorar la aproximación al EHW. En poblaciones de especies de autofecundación (autógamas) no se cumple la condición de EHW. En otras situaciones donde las poblaciones experimentales no son producto del apareamiento aleatorio sino del control deliberado del apareamiento por parte del experimentador (selección), tampoco se alcanza el EHW. De cualquier modo, considerando un locus, podrían alcanzarlo mediante una simple generación de apareamiento aleatorio. Con el conocimiento de las frecuencias génicas y en ausencia de selección natural es posible determinar cuáles serán las frecuencias genotípicas en el equilibrio.

2.3. Número o tamaño efectivo poblacional y deriva genética

El reducido tamaño de una población y la variabilidad en el número de descendientes que aportan los individuos que la componen a la siguiente generación son factores que incrementan la probabilidad de deriva genética de la población. La deriva genética es un proceso aleatorio por el cual, a partir de una frecuencia génica inicial y por efecto del muestreo imperfecto de gametas causado por un tamaño poblacional reducido, la población experimenta un incremento en la frecuencia de genotipos homocigotas en detrimento de los heterocigotas hasta que en el límite del proceso los individuos alcanzarán la homocigosis para uno u otro de los alelos inicialmente segregantes en la población. Como veremos más adelante, si existen efec-



tos de dominancia en el sentido favorable, la media de la población para el carácter considerado puede verse afectada por reducirse la proporción de individuos heterocigotas.

Una manera de cuantificar el tamaño real poblacional es efectuar un censo y determinar el número de individuos (N) que la componen. Pero por sus efectos sobre las propiedades de una población, resulta más apropiado medir el tamaño de la población a partir de su tamaño o número efectivo (N_e) (Falconer y Mackay, 2001). Casi como regla general, en la naturaleza el número real resulta diferente al número efectivo. Por su efecto sobre el número efectivo, no solo importa el número de individuos de la población, sino también el modo reproductivo de la especie, las probabilidades de autofecundación y la variabilidad en la cantidad de descendientes con los que los individuos contribuyen a la siguiente generación. Esta última condición puede responder a diferentes causas, como factores aleatorios, variabilidad ambiental y variabilidad genética presentes en la población (Gale y Lawrence, 1984). La frecuencia de autofecundaciones es la que genera la mayor reducción del N_e respecto al N . Durante el proceso de selección en poblaciones experimentales o de cría, el fitomejorador puede conocer y controlar el tamaño efectivo.

2.3.1. Cálculo del tamaño o número efectivo

2.3.1.1. Efecto del hábito reproductivo

El tamaño censal y el tamaño efectivo poblacional pueden ser diferentes en función del sistema de apareamiento de la especie (Gale y Lawrence, 1984). El modo de reproducción natural de una especie tiene una clara incidencia sobre el tamaño efectivo poblacional. Las poblaciones de plantas alógamas, especialmente las autoincompatibles, tienden a poseer mayor tamaño efectivo que las de plantas autógamas (Lin *et al.*, 2014).

En especies autoincompatibles el número efectivo resulta mayor que el número real de individuos de una población:

$$N_e = N + \frac{1}{2}$$

Dentro del grupo de especies autocompatibles podemos considerar el caso de las especies alógamas y el de las especies autógamas. En las primeras las autofecundaciones pueden ocurrir de manera aleatoria, pero en muy baja frecuencia, en cuyo caso N_e puede ser igual o levemente menor a N . De acuerdo a Gale y Lawrence (1984), el efecto de la tasa de autofecundación (s) afecta el tamaño efectivo (Figura 2.2) según la expresión:

$$N_e = N - (N/2)s = N [1 - (1/2)s]$$

En especies autóгамas, donde $s=1$ (todos los individuos se originan por autofecundación), o en especies alógamas por manipulación de la polinización:

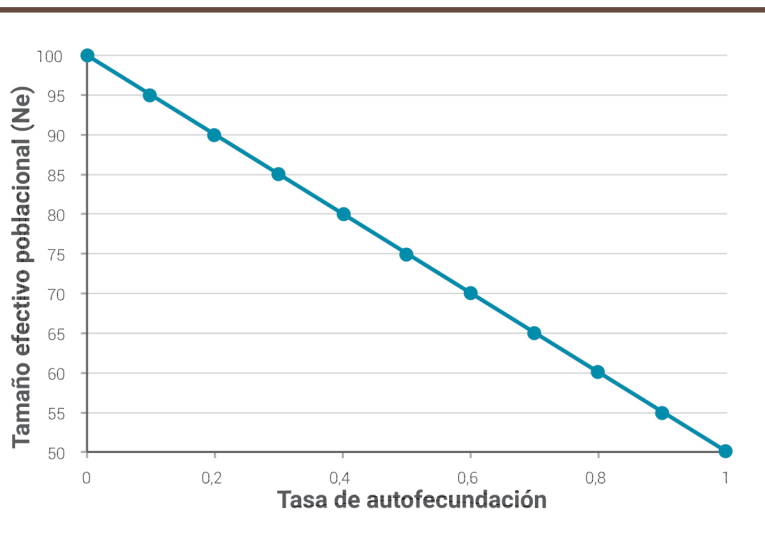


Figura 2.2. Efecto de la tasa de autofecundación en una especie diploide sobre el tamaño efectivo poblacional de una población real de tamaño 100.

$$N_e = N(1 - 1/2) = 1/2N$$

2.3.1.2. Efecto de la variabilidad en el número de descendientes por individuo

Designaremos como k al número de alelos efectivamente transmitidos por cada individuo a la siguiente generación (uno por cada uno de sus descendientes, tratándose de un organismo diploide), y como V_k a la variancia de tal número de alelos entre individuos (debida exclusivamente a causas aleatorias y ambientales). Consideraremos una situación de completa aloгамia (apareamiento aleatorio y $s=0$). La expresión para calcular el tamaño efectivo será (Gale y Lawrence, 1984):

$$N_e = (4N - 2)/(2 + V_k), \text{ donde}$$

$$V_k = 2 - 2/N$$

En estas circunstancias, resultará

$$N = N_e$$

El caso extremo se dará si $V_k = 0$, entonces:

$$N_e = 2N - 1$$

Por el contrario, el tamaño efectivo será $N_e < N$ y declinará rápidamente (Figura 2.3) cuanto mayor sea la variancia V_k .

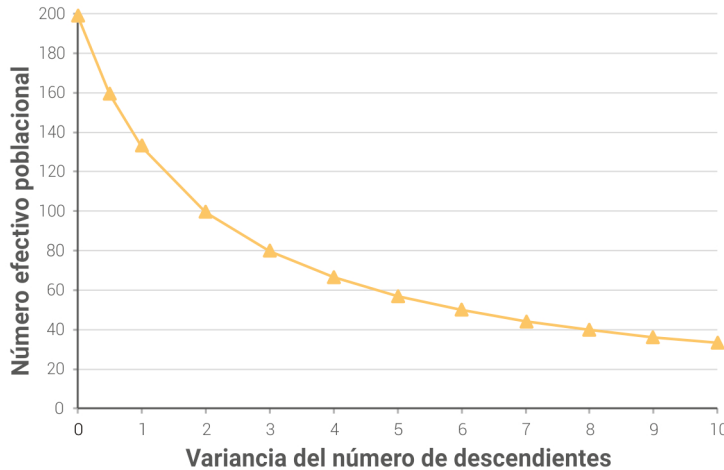


Figura 2.3. Efecto del aumento de la variancia en el número de descendientes por individuo sobre el tamaño efectivo poblacional asumiendo que el número real de individuos sea 100 y ocurra alogamia completa ($s=0$).

Si V_k no dependiera de causas aleatorias o ambientales sino genéticas (selección natural o artificial), el tamaño efectivo poblacional responde a la siguiente expresión (Nei y Murata, 1966; citados por Gale y Lawrence, 1984):

$$N_e = 4 N / [2 + (1 + 3 h^2) V_k]$$

donde

h^2 es la heredabilidad en sentido estricto (ver Capítulo 5) del carácter "número de descendientes".

Si deliberadamente provocamos que $V_k = 0$, entonces:

$$N_e = 2N$$

2.3.1.3. Efecto de la variabilidad en el tamaño efectivo poblacional a través de generaciones

Si el tamaño efectivo poblacional no fuera constante a través de t generaciones, el tamaño efectivo luego de t generaciones será la media armónica (ver Anexo) de los tamaños efectivos poblacionales en cada generación:

$$1/N_e = (1/t) [(1/N_{e1}) + (1/N_{e2}) + \dots + (1/N_{et})]$$

donde:

$N_{e1}, N_{e2}, N_{e3}, \dots, N_{et}$ son los tamaños efectivos a lo largo de las generaciones de 1 a t .

En consecuencia, la generación cuyo tamaño efectivo haya sido el menor, será la que hará la contribución relativa más importante a la disminución del tamaño efectivo final. Esto tiene implicancias prácticas en programas de mejoramiento de poblaciones, por cuanto las generaciones con menor tamaño efectivo poblacional pueden tener un efecto de cuello de botella que limite la respuesta máxima lograble por selección a largo plazo, debido a los efectos que la deriva genética pudiera ocasionar sobre la variancia genética poblacional.

Una completa revisión sobre el cálculo del tamaño efectivo de población y métodos de estimación en diferentes situaciones puede encontrarse en Caballero (1994).

2.4. Absorción de alelos y deriva genética

Si consideramos un locus con dos alelos, A y a, cada uno con frecuencias génicas poblacionales p y q , respectivamente, en una población diploide con tamaño real N , la cantidad de alelos A y a presentes en la población será $2N$. Si en la población completa existe una cantidad P de alelos A, la frecuencia génica del alelo A será:

$$p = P/2N$$

Por efecto de la deriva genética, la frecuencia del alelo A oscilará en forma aleatoria y puede producirse eventualmente su pérdida o fijación. En ausencia de selección natural y siguiendo a Wright (1931), las probabilidades de fijación y pérdida del alelo A serán $\text{Prob}(p=1) = p$ y $\text{Prob}(p=0) = (1-p) = q$, respectivamente.

Consecuentemente, cuanto más raro sea el alelo en una población (más baja su frecuencia), mayor será la probabilidad de su pérdida por deriva genética transcurridas un número de generaciones.

El término absorción de un alelo se refiere a la situación en la cual este se ha fijado o perdido en la población. Eventualmente indicaremos como generación t a aquella en la que todos los individuos de la población llegan a la homocigosis para uno u otro alelo. El tiempo de absorción t resulta ser una variable aleatoria, porque la deriva genética es un fenómeno aleatorio. Es posible estimar un tiempo medio de absorción. Notar que se considera que no existe selección natural que actúe sobre la variabilidad entre los alelos en cuestión respecto de su aptitud para perpetuarse en la población.

El tiempo medio de absorción se mide en número de generaciones, expresa la pérdida de variabilidad en determinado locus y depende de la frecuencia

alélica de partida y del tamaño efectivo de la población. Watterson (1962) y Ewens (1963) demostraron que el tiempo medio de absorción (T_M) puede aproximarse a:

$$T_M = -4 N_e [p \ln(p) + q \ln(q)]$$

Cuanto mayor sea el tamaño efectivo, mayor será el tiempo medio de absorción y menor la tasa de disminución de variabilidad genética.

2.5. Número efectivo y probabilidad de fijación por selección del alelo favorable

Crow y Kimura (1970) demostraron que la probabilidad de fijación del alelo favorable [$\text{Prob}(p=1)$] en ausencia de dominancia es:

$$\text{Prob}(p=1) = (1 - e^{-N_e s p}) / (1 - e^{-N_e s}),$$

donde

N_e es el tamaño efectivo poblacional,

s es el coeficiente de selección,

p es la frecuencia génica inicial del alelo favorable y

$e = 2,7183$.

A manera de ejemplo, aplicando esta expresión cuando $N_e=50$, $s=0,2$ y para dos casos alternativos de frecuencias génicas iniciales, intermedia ($p=0,5$) y muy baja ($p=0,05$), las probabilidades de fijación del alelo favorable por selección serán $\text{Prob}(p=1) \approx 1,000$ y $\text{Prob}(p=1) = 0,393$, respectivamente. Observemos que en el segundo caso (alelo en muy baja frecuencia) es mayor la probabilidad de que el alelo favorable se pierda por muestreo (0,95) a que se fije por selección (Vencovsky, 1980).

2.6. Desequilibrio de ligamiento y desequilibrio gamético

Cuando consideremos dos o más loci simultáneamente, la situación de desequilibrio existirá toda vez que las frecuencias gaméticas observadas en la población sean significativamente diferentes a las esperadas con base en las frecuencias alélicas en esos loci. El desequilibrio puede deberse al ligamiento físico de los loci involucrados, a un simple fenómeno de muestreo aleatorio de gametas en una población de tamaño pequeño, o a ambas circunstancias a la vez.

Si partimos de dos progenitores homocigotas para dos loci (A y B), con dos alelos cada uno (A,a; B,b), con genotipos AABB (P1) y aabb (P2), los individuos resultantes de su cruzamiento serán heterocigotas y los loci de uno de sus cromosomas llevará los correspondientes alelos favorables en ambos loci, y en el cromosoma homólogo ambos alelos desfavorables. Por lo tanto esta distribución alélica es en fase de acoplamiento o asociación. De tal manera el genotipo de la F_1 será:

$$F_1 = AB//ab$$

La filial F_2 que puede obtenerse por apareamiento aleatorio o mediante autofecundación de los individuos F_1 , indistintamente, se formará a partir de la fecundación entre gametas generadas por los individuos F_1 . En este caso, las gametas generadas por tales individuos se presentan en la Tabla 2.1.

Fase	Frecuencia Gaméticas	Gametas de tipo parental		Gametas de tipo recombinante	
		AB	Ab	Ab	aB
Acoplamiento AB//ab	Esperada	$p_A p_B$	$q_A p_B$	$p_A q_B$	$q_A q_B$
	Observada	r	s	t	u

Tabla 2.1.

Frecuencias gaméticas esperadas en ausencia de ligamiento y observables en presencia de ligamiento correspondientes a la F_1 de un cruzamiento entre dos progenitores homocigotas con alelos en fase de acoplamiento para dos loci, A y B, con dos alelos por locus (A,a; B,b). C es el coeficiente de recombinación entre los loci A y B.

Una proporción de ellas, producto de la ausencia total de *crossing over* entre ambos loci (u ocurrencia de un número par de eventos de *crossing over* si no existiera interferencia), tendrá la misma composición alélica que las gametas de los progenitores del cruzamiento. A tales gametas las denominaremos gametas de tipo parental. La ocurrencia de *crossing over* generará gametas de tipo recombinante que llevan el alelo favorable para un locus y el desfavorable para el otro. Designaremos como frecuencias gaméticas r y s a las correspondientes que llevan los alelos favorables y los alelos desfavorables en ambos loci, respectivamente, es decir, a las frecuencias gaméticas parentales. Las frecuencias t y u corresponden a gametas recombinantes, es decir, que llevan el alelo favorable en el locus A y el desfavorable en el B, o bien el alelo desfavorable en A y el favorable en B.

Si los loci A y B fueran independientes (no ligados), las frecuencias gaméticas r, s, t y u deberían ser iguales a las frecuencias esperadas, es decir, $p_A p_B = r$; $q_A q_B = s$; $p_A q_B = t$ y $q_A p_B = u$. Si, por el contrario, los loci A y B estuvieran ligados, esas igualdades no se cumplen y se verificaría un

exceso de frecuencia de las gametas de tipo parental y un defecto de igual magnitud en las gametas de tipo recombinante, con respecto las frecuencias esperadas. El apartamiento de las frecuencias gaméticas observadas y las esperadas es directamente proporcional al coeficiente de recombinación entre ambos loci, c , cuyo valor máximo es 0,5, correspondiente a la situación de ausencia de ligamiento. De tal modo podemos expresar las frecuencias observadas como:

$$r = \left(\frac{1}{2}\right)(1 - c)$$

$$s = \left(\frac{1}{2}\right)(1 - c)$$

$$t = \left(\frac{1}{2}\right) c$$

$$u = \left(\frac{1}{2}\right) c,$$

y las diferencias entre las frecuencias gaméticas esperadas en el equilibrio y las observadas como:

$$D_{AB} = p_A p_B - r$$

$$D_{ab} = q_A q_B - s$$

$$D_{Ab} = p_A q_B - t$$

$$D_{aB} = q_A p_B - u$$

El desequilibrio de ligamiento se calcula como la diferencia entre las frecuencias observadas de las gametas de tipo parental y la de las gametas de tipo no parental. De tal modo:

$$D = rs - tu$$

Alternativamente, el desequilibrio de ligamiento puede calcularse a partir de las diferencias entre las frecuencias genotípicas de individuos que son heterocigotas (Δf_{HET}) con alelos distribuidos en fase de acoplamiento y en fase de repulsión.

En este caso:

$$\Delta f_{\text{HET}} = f(AB//ab) - f(Ab//aB) = 2rs - 2ut = 2(rs - ut)$$

$$\text{y } D = \left(\frac{1}{2}\right) \Delta f_{\text{HET}} = rs - ut$$

En el caso de que hubiésemos partido de dos progenitores homocigotas con alelos distribuidos en repulsión para dos loci (A y B), con dos alelos cada uno (A,a; Bb), pero con genotipos AA bb (P1) y aaBB (P2) (genes distribuidos en repulsión), los individuos resultantes de su cruzamiento serán heterocigotas y sus loci en uno de sus cromosomas llevará los correspondientes alelos favorable (A) y desfavorable (b), y en el cromosoma homólogo los alelos desfavorable (a) y favorable (B). De tal manera el genotipo de la F₁ será:

$$F_1 = Ab//aB$$

La filial F₂ que puede obtenerse por apareamiento aleatorio o mediante autofecundación de los individuos F₁, indistintamente, se formará a partir de la fecundación entre gametas generadas por los individuos F₁. En este caso, las gametas generadas por tales individuos se presentan en la Tabla 2.2.

Fase	Frecuencia Gaméticas	Gametas de tipo parental		Gametas de tipo recombinante	
		Ab	aB	AB	ab
Repulsión Ab//aB	Esperada	$p_A q_B$	$q_A p_B$	$p_A p_B$	$q_A q_B$
	Observada	r	s	t	u

Tabla 2.2.

Frecuencias gaméticas esperadas en ausencia de ligamiento y observables en presencia de ligamiento correspondientes a la F1 de un cruzamiento entre dos progenitores homocigotas con alelos distribuidos en fase de repulsión para dos loci, A y B, con dos alelos por locus (A,a; B,b). C es el coeficiente de recombinación entre los loci A y B.

Designaremos como frecuencias gaméticas t y u a las correspondientes que llevan los alelos favorables y los alelos desfavorables en ambos loci, respectivamente, es decir, a las frecuencias gaméticas recombinantes. Las frecuencias r y s corresponden a gametas parentales, es decir, que llevan el alelo favorable en el locus A y el desfavorable en B, o bien el alelo desfavorable en A y el favorable en B.

El desequilibrio D será igual a la diferencia entre el producto de las frecuencias observadas de las gametas de tipo parental y el producto de las frecuencias observadas de las gametas de tipo recombinante.

El desequilibrio de ligamiento (D) entre genes ligados ($c < 0,5$) se reduce a la mitad en cada generación de apareamiento aleatorio de acuerdo a la siguiente expresión:

$$D_t = D_o (1-c)^t,$$

donde D_t es el desequilibrio luego de t generaciones de recombinación y D_o el desequilibrio de ligamiento inicial.

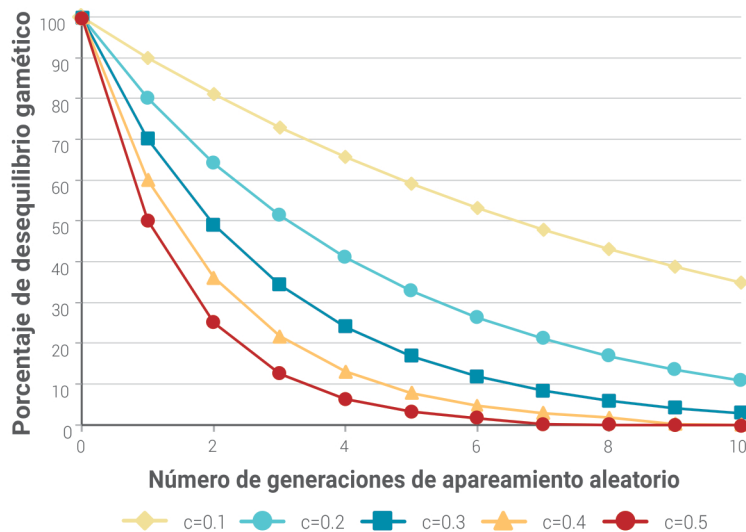


Figura 2.4. Evolución del desequilibrio gamético a lo largo de t generaciones de apareamiento aleatorio considerando diferentes coeficientes de recombinación (c) entre loci.

El desequilibrio gamético por causa del ligamiento físico entre genes es una situación muy probable cuando nos referimos a caracteres de herencia cuantitativa debido a que el número de loci involucrados excede largamente al número básico de cromosomas.

Cuanto mayor sea el ligamiento (más pequeño sea c), el desequilibrio por ligamiento persistirá por mayor número de generaciones de apareamiento aleatorio (Figura 2.4).

El desequilibrio de ligamiento también puede reducirse bajo sistemas de autofecundación puesto que la meiosis ofrece nuevas oportunidades de recombinación genética en cada generación. En especies autóгамas o con autofecundación, el desequilibrio de ligamiento puede desaparecer en una sola generación cuando los individuos son heterocigotas en ambos loci.

Cuando consideramos el ligamiento entre pares de loci, es interesante notar que la reducción en el desequilibrio de ligamiento tiene que ver con los eventos de recombinación genética que ocurren en los individuos heterocigotas para ambos loci, puesto que son los únicos que permiten la formación de nuevas clases gaméticas o de clases gaméticas que estaban en menor frecuencia respecto de la esperada. El entrecruzamiento genético o *crossing-over* que ocurre en individuos homocigotas para ambos loci, o en uno solo de ellos, resulta inefectivo en la producción de nuevos tipos gaméticos.

La existencia de ligamiento genético implica la falta de independencia entre los efectos genotípicos de los loci involucrados. Esta falta de independencia se expresa por la contribución adicional de las covariancias entre los loci a la variancia genotípica total.

El desequilibrio de ligamiento es un fenómeno que interesa a nivel de individuo, pero también tiene implicancias a nivel poblacional y está relacionado con su tamaño efectivo. Es decir que el desequilibrio de ligamiento está afectado por factores biológicos, como la tasa de recombinación entre genes (c) y por factores históricos que afectan o afectaron el tamaño efectivo poblacional (N_e). Según Gaut y Long (2003), el desequilibrio de ligamiento esperado [$\mathcal{E}(D^2)$] en una población puede expresarse de la siguiente manera:

$$\mathcal{E}(D^2) = 1/(1 + 4 c N_e),$$

donde c es la tasa de recombinación entre loci (o entre dos marcadores) medida en Morgans. En consecuencia la condición en la que podrá esperarse mayor desequilibrio de ligamiento será cuanto más próximos se encuentren los genes (o bien un marcador molecular respecto a su QTL más cercano) sobre el cromosoma y cuanto menor sea el tamaño efectivo poblacional.

Los eventos y las características de una población que afectan al tamaño efectivo poblacional y con ello al desequilibrio de ligamiento comprenden:

- El nivel de endocria
- La aplicación de selección direccional
- La existencia de cuellos de botella en el tamaño de la población
- La estructura de la población (número de subpoblaciones, ocurrencia de mezclas, procesos de migración)

Estos cuatro factores disminuyen el tamaño efectivo poblacional y por lo tanto aumentan el desequilibrio de ligamiento esperado. El sistema de apareamiento de la especie también afecta al desequilibrio de ligamiento. En especies autóгамas, en comparación con las alógamas, para una misma distancia física en el cromosoma medida en miles de pares de bases ocurre mayor desequilibrio de ligamiento (Wright *et al.*, 2008; Gupta *et al.*, 2005), o dicho de otra manera, para un determinado nivel de desequilibrio de ligamiento la longitud de los segmentos cromosómicos es mayor en especies autóгамas que alógamas. Ello puede deberse a que en el caso de las especies autóгамas los eventos de entrecruzamiento no resultan efectivos para el surgimiento de nuevas combinaciones gaméticas debido a la alta homocigosis a través de loci. Por el contrario, en especies de polinización cruzada, la mayor heterocigosis contribuye a la formación de nuevos tipos gaméticos como resultado del entrecruzamiento. El número y longitud de los segmentos dentro de los cuales hay equilibrio de ligamiento es un factor muy importante a tener en cuenta en programas de selección genómica, por cuanto el número mínimo de marcadores moleculares debe ser el necesario para asegurar que exista al menos uno de ellos en cada uno de aquellos segmentos (ver con más detalle en el Capítulo 10).

2.7. Variabilidad genética

La variabilidad de origen genético presente en una población para un carácter cuantitativo depende del número de loci, de la diversidad alélica y de los modos de acción génica. No revisaremos en este capítulo los mecanismos que generan tal variabilidad, sino que describiremos cómo los dos primeros factores determinan la existencia de individuos genéticamente diferentes y más adelante nos ocuparemos de cuantificar y describir esta variabilidad a nivel poblacional.

2.7.1. Clases genotípicas y número de genotipos cuando hay dos alelos por locus

Es importante aclarar que las expresiones que se presentan en esta sección se refieren al cálculo del número de clases genotípicas, diferenciándola del número de genotipos posibles. Estamos asumiendo que los loci cuantitativos tienen efectos similares e intercambiables y por lo tanto desde un punto de vista práctico cada clase genotípica se corresponderá con una clase fenotípica (asumiendo ausencia de efectos ambientales o efectos similares del ambiente sobre cada genotipo).

Considerando dos alelos por locus, el número de clases genotípicas posibles (${}^mN_{CG}$) es función del número de loci (m):

$${}^mN_{CG} = 2m + 1,$$

A partir del teorema del binomio, podemos describir al número de clases genotípicas considerando 1, 2, 3... m loci de la siguiente manera:

Para un locus ($m = 1$):

$$(p+q)^2 = p^2 + 2pq + q^2,$$

donde p representa la frecuencia del alelo favorable y q la del desfavorable. Es decir, que habrá tres clases genotípicas, una con individuos en una frecuencia p^2 con los dos alelos favorables, otra con individuos con los dos alelos desfavorables en una frecuencia q^2 y la tercera con individuos en una frecuencia $2pq$ que llevan un alelo favorable y otro desfavorable.

El número de genotipos posibles para $m = 1$ coincidirá con el número de clases genotípicas:

$${}^mN_G = 3^m$$

Para dos loci el número de genotipos posibles será:

$${}^2N_G = 3^2 = 9$$

y el número de clases genotípicas (o fenotípicas):

$${}^2N_{CG} = (2)(2) + 1 = 5$$

$(p+q)^{2m} = (p+q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$, es decir, que habrá cinco clases genotípicas con las siguientes características: una con individuos con los dos alelos favorables en los dos loci (con frecuencia p^4), otra con individuos con los dos alelos desfavorables en los dos loci (q^4), otra con individuos que poseen tres alelos favorables en total y uno desfavorable, independientemente de su ubicación en los loci ($4p^3q$), otra con individuos con un alelo favorable y tres desfavorables ($4pq^3$) y otra con individuos con dos alelos favorables y dos desfavorables ($6p^2q^2$), prescindiendo de cómo se distribuyan en los loci.

En estas expresiones y para simplificar, asumimos que los dos loci en la población poseen la misma frecuencia del alelo favorable y la misma frecuencia del alelo desfavorable. Observemos que los términos de desarrollo del binomio corresponden a productos entre potencias de las dos frecuencias génicas, precedidos por un coeficiente. Los términos comienzan con la frecuencia génica del alelo favorable elevada a la potencia $2m$, donde m es el número de loci, el segundo término con la misma frecuencia elevado a la potencia $2m-1$ multiplicado por la frecuencia del alelo desfavorable a la potencia 1, el siguiente término por la frecuencia del alelo favorable a la potencia $2m-2$ multiplicado por la frecuencia del alelo desfavorable a la potencia 2 y así sucesivamente hasta terminar por la frecuencia del alelo desfavorable a la potencia $2m$. Los exponentes indican el número de alelos favorables o desfavorables para cada clase genotípica.

Los coeficientes que acompañan a cada término pueden obtenerse, alternativamente, a partir de:

$$C_{2m} = (2m)! / f! d!,$$

donde

$(2m)!$ es el factorial del doble del número de loci,

$f!$ es el factorial del número de alelos favorables y

$d!$ es el factorial del número de alelos desfavorables.

Por ejemplo, para 3 loci, el número de individuos que portan 4 favorables y 2 desfavorables (p^4q^2) será:

$$C_6 = 6! / 4! 2! = 6 \times 5 \times 4! / 4! \times 2 \times 1 = 15$$

Si hiciéramos los cálculos para un número mayor de loci, podríamos comprobar que los coeficientes o frecuencias de cada clase genotípica se acomodan conforme el Triángulo de Tartaglia o de Pascal (Figura 2.5).

				1							
$(p+q)^{2 \times 1}$			1	2	1						
			1	3	3	1					
$(p+q)^{2 \times 2}$			1	4	6	4	1				
			1	5	10	10	5	1			
$(p+q)^{2 \times 3}$			1	6	15	20	15	6	1		
			1	7	21	35	35	21	7	1	
$(p+q)^{2 \times 4}$			1	8	28	56	70	56	28	8	1

Figura 2.5.
Triángulo de Tartaglia aplicado al cálculo del número y frecuencia relativa de clases genotípicas correspondientes a un organismo diploide, con dos alelos por locus y número variable de loci.

2.7.2. Número de clases genotípicas posibles cuando hay múltiples alelos por locus

En los casos de alelos múltiples ($n > 2$), el número de clases genotípicas posibles es el número de términos del cuadrado de un polinomio de orden n . Para el caso de un locus con tres alelos:

$${}^nN_{CG} = (p+q+r)^2 = (p+q+r)(p+q+r) = p^2 + 2pq + 2pr + q^2 + 2rq + r^2,$$

es decir, que el número de clases genotípicas posibles es seis (Tabla 2.3.)

Cuando hay cuatro alelos por locus:

$${}^nN_{CG} = (p+q+r+s)^2 = (p+q+r+s)(p+q+r+s) = p^2 + 2pq + 2pr + 2ps + q^2 + 2rq + r^2 + 2sp + 2sq + 2sr + s^2$$

el número de clases genotípicas posibles es 10.

Para cinco alelos por locus:

$${}^nN_{CG} = (p+q+r+s+t)^2 = (p+q+r+s+t)(p+q+r+s+t) = p^2 + 2pq + 2pr + q^2 + 2rq + r^2 + 2sp + 2sq + 2sr + s^2 + 2tp + 2tq + 2tr + 2ts + t^2,$$

por lo tanto aquel número es 15.

En concordancia con ello, en una especie diploide la expresión correspondiente al número de clases genotípicas por cada locus segregante (CR^2_n) es el número combinatorio con repeticiones de n elementos (alelos) tomados de a dos:

$$CR^2_n = {}^nN_G = (n+2-1)! / 2! (n-1)!$$

Si el carácter estuviera controlado por un número variable de loci (m), cada uno con n alelos múltiples, la expresión debe modificarse de la siguiente manera:

$${}^mNG_n = [(n+2-1)! / 2! (n-1)!]^m \quad (\text{Tabla 2.4})$$

Número de alelos (n) por locus	Número de clases genotípicas posibles
2	3
3	6
4	10
5	15
6	21
10	55
n	$(n+2-1)! / 2! (n-1)!$

Tabla 2.3.

Número de clases genotípicas posibles para un locus en función del número de alelos por locus.

Número de alelos por locus (n)	Número de loci (m)			
	1	2	10	50
2	3	9	59,049	7,18 x 10 ²³
3	6	36	60.466.176	8,08 x 10 ³⁸
4	10	100	10.000.000.000	1,00 x 10 ⁴¹
5	15	225	576.650.390.600	6,37 x 10 ⁵⁸
6	21	441	16.679.880.980.000	1,29 x 10 ⁶⁶
10	55	3025	2,53 x 10 ¹⁷	∞
n	[(n+2-1)! / 2!(n-1)!] ^m			

Tabla 2.4.

Número de clases genotípicas posibles en función del número de alelos por locus (n) y número de loci (m) en una especie diploide.

2.7.3. Número de clases fenotípicas

Si consideramos solo dos alelos por locus y dominancia parcial, la cantidad de fenotipos posibles se calcula con la expresión 3^m , donde m es el número de loci considerados.

Cuando la dominancia es completa, solo dos fenotipos de tres genotipos podrán ser distinguibles y por lo tanto la cantidad de clases fenotípicas será 2^m donde m es el número de loci involucrados. Suponiendo que del total de m loci hay (m-x) que tienen dominancia completa y x loci tienen dominancia parcial y por lo tanto hacen distinguibles los tres genotipos posibles, la cantidad de clases fenotípicas será:

$$N_p = 3^x 2^{m-x}$$

2.7.4. Proporción de individuos heterocigotas y homocigotas

Las proporciones de individuos heterocigotas y homocigotas para un número variable de alelos por locus siguen las siguientes expresiones:

$$\text{Fracción de homocigotas } (P_{ii}) = 1/n$$

$$\text{Fracción de heterocigotas } (P_{ij}) = 1 - 1/n,$$

donde n es el número de alelos por locus (ver Tabla 2.5).

Número de alelos por locus (n)	Fracción heterocigota	Fracción homocigota
	P_{ij}	P_{ii}
2	1/2	1/2
3	2/3	1/3
4	3/4	1/4
5	4/5	1/5
6	5/6	1/6
n	$1 - 1/n$	1/n

Tabla 2.5. Frecuencias relativas de individuos homocigotas y heterocigotas en una población para un locus y un número variable de alelos por locus en una especie diploide

En el caso de m loci controlando el carácter, las proporciones de heterocigotas (P_{ij}) y homocigotas (P_{ii}) serán, respectivamente (Tabla 2.6):

Número de alelos por locus (n)	Número de loci controlando el carácter (m)					
	1		2		3	
	P_{ij}	P_{ii}	P_{ij}	P_{ii}	P_{ij}	P_{ii}
2	1/2	1/2	3/4	1/4	7/8	1/8
4	3/4	1/4	15/16	1/16	63/64	1/64
6	5/6	1/6	35/36	1/36	215/216	1/216
n	$1-(1/n)^1$	$(1/n)^1$	$1-(1/n)^2$	$(1/n)^2$	$1-(1/n)^3$	$(1/n)^3$

Tabla 2.6. Frecuencias relativas de individuos homocigotas y heterocigotas para m loci y un número variable de n alelos por locus en una especie diploide.

Fracción de heterocigotas: ${}^mP_{ij} = 1-(1/n)^m$

Fracción de homocigotas: ${}^mP_{ii} = (1/n)^m$

Para un genotipo en particular es posible calcular su frecuencia en función de las frecuencias genotípicas en cada uno de los loci que lo componen, asumiendo que los loci son independientes. A modo de ejemplo, en una población en equilibrio de HW donde las frecuencias alélicas para tres loci independientes G, H y J son p, q para el locus G; r y s para el H; y t, u para el J, pueden calcularse las frecuencias genotípicas a partir del producto de las frecuencias de equilibrio para cada locus. Así por ejemplo,

$$GG \text{ HH JJ: } (p^2) \times (r^2) \times (t^2)$$

$$Gg \text{ HH Jj: } (2pq) \times (r^2) \times (2tu)$$

$$Gg \text{ hh JJ: } (2pq) \times (s^2) \times (t^2)$$

En el caso de tratarse de una población con frecuencias génicas de 0,5 en los tres loci, las frecuencias serán:

$$GG \text{ HH JJ: } (1/4)(1/4)(1/4) = 1/64 = (1/4)^m = (1/4)^3$$

$$Gg \text{ hh Jj: } (1/2)(1/4)(1/2) = 1/16 = (1/2)^{m'}(1/4)^m = (1/2)^2(1/4)$$

$$Gg \text{ hh JJ: } (1/2)(1/4)(1/4) = 1/32 = (1/2)^{m'}(1/4)^m = (1/2)(1/4)^2,$$

donde m es el número de loci al estado homocigota y m' el número de loci heterocigotas.

Las proporciones de individuos heterocigotas y homocigotas que se rigen por las expresiones descritas más arriba se refieren a poblaciones en EHW. En el caso de sistemas de endocría, por ejemplo por autofecundación, las proporciones se van modificando en función del número de generaciones de autofecundación, incrementándose en forma relativa la frecuencia de genotipos homocigotas a expensas de los heterocigotas.

Supongamos que partimos del cruzamiento entre dos progenitores homocigotas y que ambos son polimórficos para un locus. Aunque existan alelos múltiples por locus, tratándose de la F_1 de una especie diploide, habrá en juego siempre solo dos variantes alélicas, una por cada progenitor. En tal caso la constitución genotípica de la F_1 , de la F_2 , de la F_3 y generaciones posteriores hasta completar s generaciones de autofecundación desde la F_1 se calculan mediante las expresiones siguientes:

$$\text{Frecuencia de cada homocigota: } [1 - (1/2)^s]/2$$

$$\text{Frecuencia de los heterocigotas: } (1/2)^s$$

Alternativamente pueden utilizarse las expresiones siguientes:

$$\text{Frecuencia de cada homocigota: } [1 - (1/2)^{g-1}]/2$$

$$\text{Frecuencia de los heterocigotas: } (1/2)^{g-1},$$

donde g indica la filial de segregación (ej. para F₅: s = 4; g = 5)

La Tabla 2.7 ilustra las frecuencias correspondientes.

Generación	HH	Hh	hh	P _{ii}
P ₁	1	0	0	1
P ₂	0	0	1	1
F ₁	0	1	0	0,0000
F ₂	¼	½	¼	0,5000
F ₃	3/8	¼	3/8	0,7500
F ₄	7/16	1/8	7/16	0,8750
F ₅	15/32	1/16	15/32	0,9375
F ₆	31/64	1/32	31/64	0,9688
F ₇	63/128	1/64	63/128	0,9844
F ₈	127/256	1/128	127/256	0,9922
F _g	$[1-(1/2)^{g-1}]/2$	$(1/2)^{g-1}$	$[1-(1/2)^{g-1}]/2$	$[2^{g-1}-1]/2^{g-1}$

Tabla 2.7.

Frecuencia de clases genotípicas y proporción de homocigotas (P_{ii}) en generaciones derivadas por autofecundación a partir de la F₁ entre dos progenitores homocigotas y polimórficos para el locus H.

Nótese que la frecuencia de endocría es la frecuencia de individuos homocigotas idénticos por descendencia en cada generación, es decir,

$$P_{ii} = [1-(1/2)^s] = [1-(1/2)^{g-1}] = (2^s - 1)/2^s \quad [2.1]$$

En el caso de m loci o factores segregantes, la frecuencia de homocigotas en cada filial puede calcularse también a partir de la siguiente expresión (Hallauer, 1992):

$$P_{ii} = [(2^s - 1) / 2^s]^m \quad [2.2]$$

donde

s es el número de las generaciones de autofecundación para contar después de la F_1 y m el número de factores segregantes.

Así por ejemplo, la proporción de homocigotas en la filial F_6 para un locus segregante será, aplicando [2.1]:

$$P_{ii} = [(2^s - 1) / 2^s] = [(2^5 - 1) / 2^5] = 31/32 = 0,9688$$

Para la F_6 y tres loci segregantes, aplicando la expresión [2.2]:

$$P_{ii} = [(2^5 - 1) / 2^5]^3 = 0,9091$$

Si las dos líneas progenitoras resultaran polimórficas para más de un locus y siempre que estos sean independientes, podríamos calcular las frecuencias de las clases genotípicas de manera análoga a la presentada más arriba y en la Tabla 2.5.

Así, por ejemplo, la frecuencia esperada de individuos de diversas clases genotípicas en la filial F_3 para los loci G-, H- y J- serán:

$$GGHHJJ: (3/8)(3/8)(3/8) = 27/512 = 0,0527$$

$$GgHhJj: (1/4)(1/4)(3/8) = 3/128 = 0,0234$$

$$gghhjj: (3/8)(3/8)(1/4) = 9/256 = 0,0352$$

En la filial F_5 serán:

$$GGHHJJ: (15/32)(15/32)(15/32) = 3375/32768 = 0,1030$$

$$GgHhJj: (1/16)(1/16)(15/32) = 15/8192 = 0,0018$$

$$gghhjj: (15/32)(15/32)(1/16) = 225/16384 = 0,0137$$

2.8. Consideraciones sobre muestreo poblacional Probabilidad de encontrar un genotipo deseado

Cuanto mayor sea el número de loci que controla al carácter de interés, menor será la frecuencia esperada de un genotipo en particular. Tanto en el trabajo de mejoramiento como en el de conservación de la variabilidad de los recursos genéticos, es extremadamente importante la calidad de la muestra poblacional que asegure la inclusión en ella de los genotipos deseados, en el primer caso, y de todos los alelos existentes y no perder variabilidad genética en el segundo.

Considerando una especie diploide y un carácter controlado por un número m de loci, todos ellos con frecuencias p para el alelo favorable y q para el desfavorable, es posible conocer las diferentes clases genotípicas y describirlas en términos del número de alelos favorables que poseen.

La selección se practica sobre muestras poblacionales, cuya representatividad puede condicionar el éxito de un proyecto de selección. Para ello resulta imprescindible que la muestra incluya los alelos favorables necesarios al cumplimiento de los objetivos de mejoramiento.

Asumamos que en una población P de tamaño real N existe una cantidad de individuos N_f con la constitución genotípica deseada y una cantidad $N_{nf} = N - N_f$ con los genotipos no deseados. Será importante conocer la probabilidad de que una muestra de n individuos tomada de P incluya k individuos deseados y $n-k$ individuos no deseados. Esta probabilidad puede seguir diferente distribución según cual sea el tamaño poblacional (Crossa, 1989). Si el tamaño (N) de la población P es pequeño, el número k de individuos o genotipos deseados incluidos en muestras de distintos tamaños (n) sigue una distribución hipergeométrica:

$$N_{(k)} = \frac{N_f! (N-n)!}{N! (N_f - n)!}$$

El número esperado de individuos k en una muestra de tamaño n se puede calcular como:

$$\varepsilon(k) = n(N_f/N)$$

Cuando la población P es de tamaño grande la distribución del número de individuos favorables (k) sigue una distribución binomial. La frecuencia de individuos no deseados y deseados en la población serán, respectivamente, $A = N_{nf}/N$ y $B = N_f/N$. Por cada individuo que se agregue a la muestra, hay una probabilidad A de obtener un genotipo no deseado. Por lo tanto la probabilidad de haber incluido al menos un genotipo deseado en una muestra de tamaño n será (Mainland, 1923):

$$P = 1 - A^n$$

Por lo tanto $n = \lceil \log(1-P) / \log(A) \rceil$

Esta última expresión permite ilustrar la importancia del tamaño de muestra cuando la frecuencia del genotipo favorable es muy baja (Tabla 2.8).

Probabilidad (P)	Tamaño de muestra (n) Proporción del genotipo raro (B) = p ²				
	0,01	0,03	0,05	0,07	0,11
0,90	229	76	45	32	20
0,93	265	87	52	37	23
0,96	320	106	63	44	28
0,99	458	151	90	63	40

Tabla 2.8. Tamaños de muestras poblacionales requeridos para tener una probabilidad (P) de que la muestra haya incluido al menos un individuo deseado homocigota para el alelo favorable cuando estos se encuentran a baja frecuencia (B).

Cuanto menor sea la frecuencia del genotipo favorable, mayor será el tamaño de muestra necesario para tener una determinada probabilidad de haber incluido en la muestra al menos un individuo con el genotipo deseado.

Muller (1923), citado por Hallauer (1992) y Francis *et al.* (2012), desarrolló una fórmula para calcular la mínima cantidad de plantas que deben muestrearse de una población derivada de la filial F₁ entre líneas homocigotas para disponer de al menos una con el genotipo deseado con determinado nivel de probabilidad:

$$K = \ln(1-P) / \ln(1-B) \quad [2.3],$$

donde

K=mínimo tamaño muestral requerido

P= probabilidad de éxito deseada

B= frecuencia del genotipo deseado en la población.

La aplicación de la fórmula se presenta en la Tabla 2.9 (a partir de Hallauer, 1992) aplicadas a varias generaciones de autofecundaciones [a] y de retrocruzas [b].

[a]				
Número de loci segregantes (m)	Frecuencia del genotipo deseado en la población resultante de autofecundación	Niveles de probabilidad		
		0,90	0,99	
1	1/4	8	16	
2	1/16	36	71	
3	1/64	146	292	
4	1/256	588	1175	
5	1/1024	2354	4708	
6	1/4096	9420	18839	

[b]				
Número de loci segregantes (m)	Frecuencia del genotipo deseado en la población resultante de retrocruzamiento	Niveles de probabilidad		
		0,90	0,99	
1	1/2	4	7	
2	1/4	8	17	
3	1/8	18	35	
4	1/16	36	72	
5	1/32	73	145	
6	1/64	146	293	

Tabla 2.9. Tamaños de muestra requeridos para disponer con cierta probabilidad de al menos un individuo homocigota favorable en función del número de loci que controlan el carácter y diferentes frecuencias del genotipo deseado en la población, resultante de autofecundaciones [a] y de retrocruzamientos [b]. Adaptado de Hallauer (1992).

Puede concluirse que la necesidad de muestrear extensivamente una única población cuando el número de factores segregantes es significativo resultará poco menos que imposible de satisfacer. Por ello para una misma cantidad de recursos, convendría muestrear más de una población y optar por un muestreo menor de individuos en un número mayor de poblaciones, buscando en cada una de ellas los genotipos que más se acerquen al óptimo sin llegar a ser homocigotas para los alelos deseados en cada uno de los m loci. Una aplicación práctica de esto en programas de mejoramiento es el reciclaje de líneas élite o el uso de poblaciones derivadas de cruza- mientos élite x élite, donde la frecuencia del genotipo “óptimo” es en cada ciclo mayor y en consecuencia las exigencias del muestreo van disminu- yendo conforme la ecuación de Muller [2.3].

En tal caso, si se muestrean un número P de poblaciones, cada una con una probabilidad T de obtener al menos una planta con el genotipo deseado, po- demos expresar la probabilidad de no encontrar al menos una planta con el genotipo deseado cuando se muestrean ese número de poblaciones como:

$$Q = (1 - T_i)^P \quad [2.4]$$

Cuanto mayor sea el número de poblaciones, mayor será la probabilidad de encontrar al menos un genotipo deseado en al menos una de ellas. En el caso de que las poblaciones ofrezcan diferentes probabilidades de obtener el genotipo deseado, la expresión debe modificarse del siguiente modo:

Siendo π la multiplicatoria de la probabilidad de obtención de al menos un genotipo deseado en la i -ésima población, donde i varía desde 1 hasta P .

$$Q = \prod (1 - T_i)$$

La existencia de ligamiento genético entre loci puede provocar desvíos en- tre las frecuencias genotípicas relativas observadas y las esperadas, en tanto que las interacciones epistáticas (ver más adelante) pueden modifi- car las medias esperadas en cualquier generación.

2.8.1. Frecuencia de individuos con el genotipo óptimo en una población en EHW

Una vez definido cual es el genotipo óptimo, si tuviésemos un dispositivo experimental que permitiera utilizar un tamaño muestral lo suficientemente grande como para que todas las clases genotípicas estuvieran incluidas en las frecuencias esperadas, podríamos calcular su frecuencia relativa. La Tabla 2.10 presenta ejemplos de cálculo de frecuencias de los genotipos homocigotas deseados en poblaciones bajo apareamiento aleatorio, en po-

blaciones endocriadas y en generaciones derivadas por autofecundación del cruzamiento entre padres homocigotas que difieren en tres loci. Se observa que cuanto menor sea la frecuencia del alelo favorable, menor será su frecuencia genotípica y que la endocria permite aumentar ligeramente su frecuencia.

	Frec. p,q; r,s; t,u	Probabilidad del genotipo óptimo		
		GGHHJJ	GGhhJJ	ggHHJJ
	p, r, t=0,1	$(0,1^2)^3 = 1 \times 10^{-6}$	$[(0,1)^2]^2(0,9)^2 = 8,1 \times 10^{-5}$	$(0,9)^2[(0,1)^2]^2 = 8,1 \times 10^{-4}$
Población F=0	p,r,t=0,5	$(0,5^2)^3 = 0,01563$	$(0,5^2)^3 = 0,01563$	$(0,5^2)^3 = 0,01563$
	p,r,t=0,75	$(0,75^2)^3 = 0,4219$	$[(0,75)^2]^2(0,25)^2 = 0,01978$	$(0,25)^2[(0,75)^2]^2 = 0,0198$
	p,r,t = 0,1	$\frac{[(0,10)^2 + (0,10)(0,9)(0,5)]^3}{(0,9)(0,5)^3}$	$\{[(0,10)^2 + (0,10)(0,9)(0,5)]^2 [(0,9)^2 + (0,9)(0,10)(0,5)]\}$	$\frac{[(0,9)^2 + (0,9)(0,1)(0,5)][(0,10)^2 + (0,10)(0,9)(0,5)]^2}{(0,9)(0,5)^2}$
		= 0,00017	0,0026	= 0,0026
Población F=0,5	p,r,t=0,5	$\frac{\{[(0,50)^2 + (0,50)^2(0,50)]^3\}}{(0,50)^2(0,50)^3}$	$\{[(0,50)^2 + (0,50)^2(0,50)]^3\}$	$\{[(0,50)^2 + (0,50)^2(0,50)]^3\}$
		= 0,0527	=0,0527	=0,0527
	p,r,t =0,75	$\frac{\{[(0,75)^2 + (0,75)(0,25)(0,50)]^3\}}{(0,25)(0,50)^3}$	$\frac{\{[(0,75)^2 + (0,75)(0,25)(0,5)]^2 [(0,25)^2 + (0,25)(0,75)(0,5)]\}}{(0,25)(0,75)(0,5)}$	$\frac{[(0,25)^2 + (0,25)(0,75)(0,5)][(0,75)^2 + (0,75)(0,25)(0,5)]^2}{(0,25)(0,5)^2}$
		= 0,2826	=0,0673	= 0,0673
F ₂	p,r,t=0,5		$(1/4)^3 = 0,01563$	
F ₅	p,r,t=0,5		$(15/32)^3 = 0,1030$	

Tabla 2.10.

Probabilidad de ocurrencia de un genotipo óptimo hipotético para tres loci independientes en diferentes tipos de poblaciones, con diferentes niveles de endogamia y frecuencias génicas (p,q en el locus G; r,s en el locus H; t,u en el locus J).

En la práctica es imposible disponer de la totalidad de la población y debemos manejarnos con muestras lo más representativas posibles. Podemos calcular la probabilidad de que se incluya en la muestra al menos un individuo con el genotipo óptimo. Para calcularlo debemos tener en cuenta dos factores: i) la frecuencia esperada del genotipo óptimo en la población y ii) el tamaño de la muestra empleado. Intuitivamente resulta claro que deberían precisarse mayores tamaños de muestra cuanto menos abundante o frecuente sea el genotipo deseado. En otras palabras, cuanto más raro resulte el genotipo, mayor tamaño debería tener la muestra.

Suponiendo que el genotipo ideal es GGhhJJ y queremos determinar su frecuencia en la filial F_5 a partir del cruzamiento de dos progenitores polimórficos para los loci G, H y J, de acuerdo a lo precedente será $(15/32)^3$, es decir, 0,1030 (Tabla 2.7).

Si tomamos una muestra de tamaño 2 de la filial F_5 , la probabilidad de no encontrar ninguna planta de genotipo GGhhJJ será:

$$Q = (1 - 0,1030) \times (1 - 0,1030) = 0,8046$$

La probabilidad de no encontrar una planta en una muestra de tamaño 10 será:

$$Q = (1 - 0,1030)^{10} = 0,3372;$$

de no encontrar al menos una planta en una muestra de tamaño 50 será:

$$Q = (1 - 0,1030)^{50} = 0,0043;$$

de no encontrar al menos una planta en una muestra de tamaño 100 será:

$$Q = (1 - 0,1030)^{100} = 0,00002$$

Lógicamente conforme aumenta el tamaño de la muestra va disminuyendo exponencialmente la probabilidad de no haber incorporado en la muestra al genotipo deseado (Q) y va aumentando la probabilidad de haber incluido al genotipo deseado (P).

En una población en equilibrio de Hardy-Weinberg es posible estimar el número de plantas que poseen al menos un alelo favorable en m loci que controlen el carácter de interés $[(p^2 + 2pq)^m]$ o que son homocigotas $[p^{2m}]$ y a partir de allí estimar el número esperado de tales plantas para muestras de 1000 y 10000 plantas, respectivamente, tomadas al azar de la población. Los números esperados se presentan en la Figura 2.6, asumiendo las mismas frecuencias de alelos favorables en cada loci.

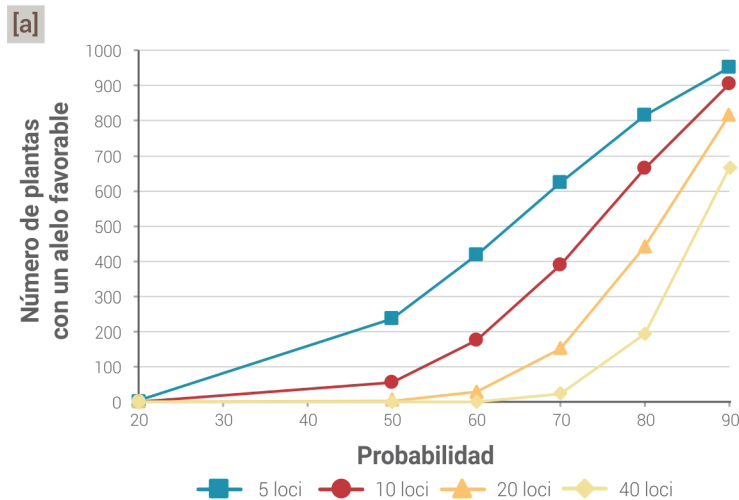
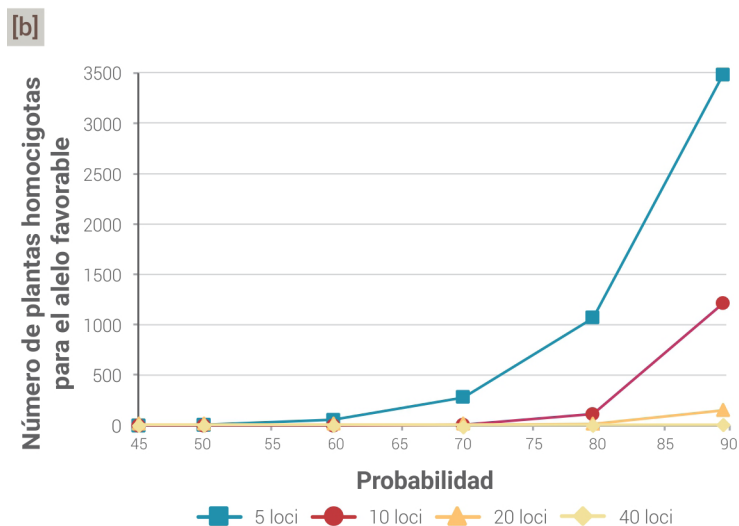



Figura 2.6. Número de plantas con un alelo favorable [a] y número de plantas homocigotas [b] para el alelo favorable en m loci que controlan un carácter en muestras de 1000 y 10000 individuos, respectivamente. A partir de Hallauer, 1987.



2.9. Determinación del tamaño de muestra adecuado en función de la frecuencia de individuos con el genotipo deseado

Sedcole (1977) abordó el problema de calcular la cantidad de individuos para muestrear una generación y tener una probabilidad determinada de haber incluido en la muestra un número de individuos que llevan los alelos deseados en uno o más loci. El tamaño de muestra (N individuos) necesario para tener K plantas con el genotipo deseado, en una población donde la frecuencia de esos genotipos es Q , dependerá de la probabilidad P o

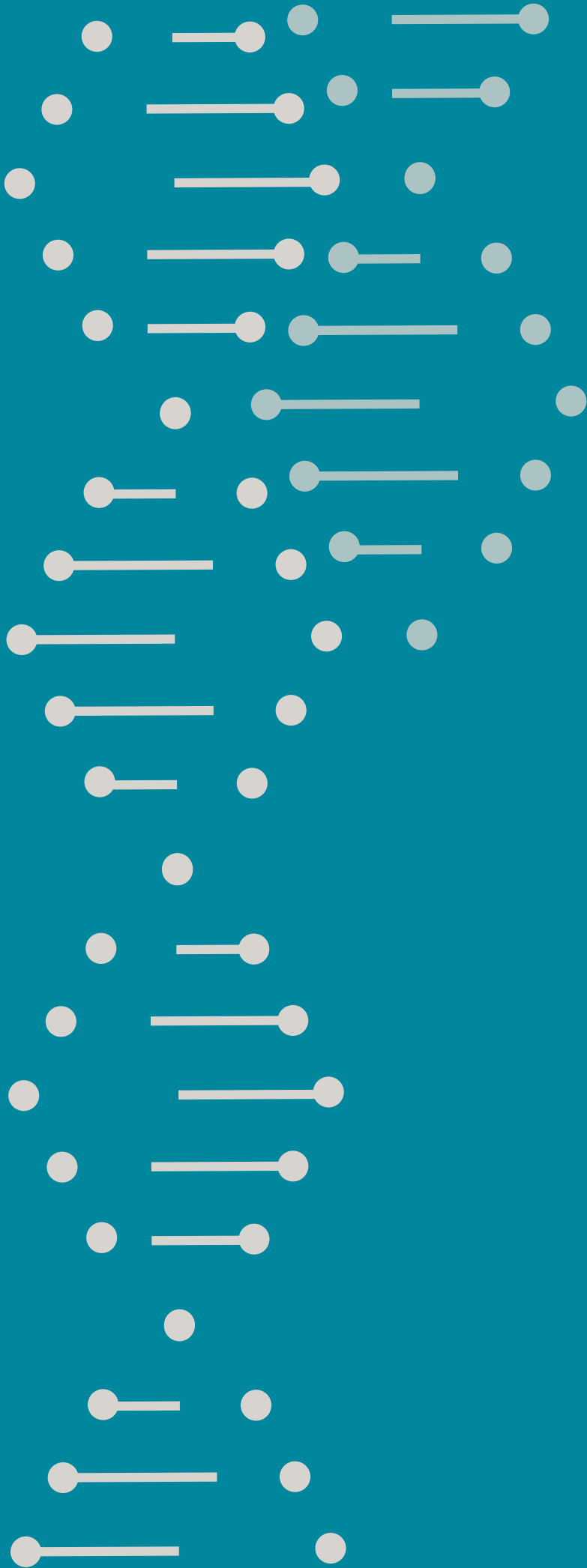


margen de seguridad con el que se desea trabajar, y z es el valor que corresponde en la curva de distribución normal para la probabilidad P .

La expresión para calcular N es:

$$N = \{2(K-0,5) + z^2(1-Q) + z[z^2(1-Q)^2 + 4(1-Q)(K-0,5)]^{0,5}\} / 2Q$$

Lógicamente el tamaño de muestra requerido será mayor cuanto mayor sea el número de plantas con el genotipo deseado que se pretende recuperar, cuanto más pequeña sea su frecuencia en la población y cuanto mayor sea la probabilidad de que así ocurra.



Modelos genéticos

3.1. Modelo genético básico

Se denomina fenotipo de un individuo a la expresión visible, calificable o cuantificable de un carácter. El genotipo de ese individuo hace referencia a su constitución genética que determina ese carácter. Como se indicó en el capítulo anterior, para los caracteres cuantitativos no existe una relación equivalente entre fenotipo y genotipo por la presencia de una desviación ambiental o simplemente el efecto del ambiente. Este comprende todos los efectos de factores que influyen o modulan la expresión de los genes que controlan el carácter, así como también al error experimental. Este concepto expresado en términos de un modelo lineal simple toma la forma:

$$P=G+E, \quad [3.1]$$

donde P es el fenotipo, G el genotipo y E la desviación ambiental.

En los caracteres cuantitativos, cuya distribución de frecuencias fenotípicas generalmente se ajusta a la distribución normal o de Gauss, idealmente podríamos visualizar la distribución probabilística que corresponda a los componentes del fenotipo, esto es la distribución de los efectos genotípicos y la de los desvíos ambientales. Como puede verse en la Figura 3.1, dependiendo de cuanto difieran esas distribuciones en sus parámetros de posición y dispersión, mayor puede ser la diferencia entre el valor fenotípico (P) y el genotípico (G).

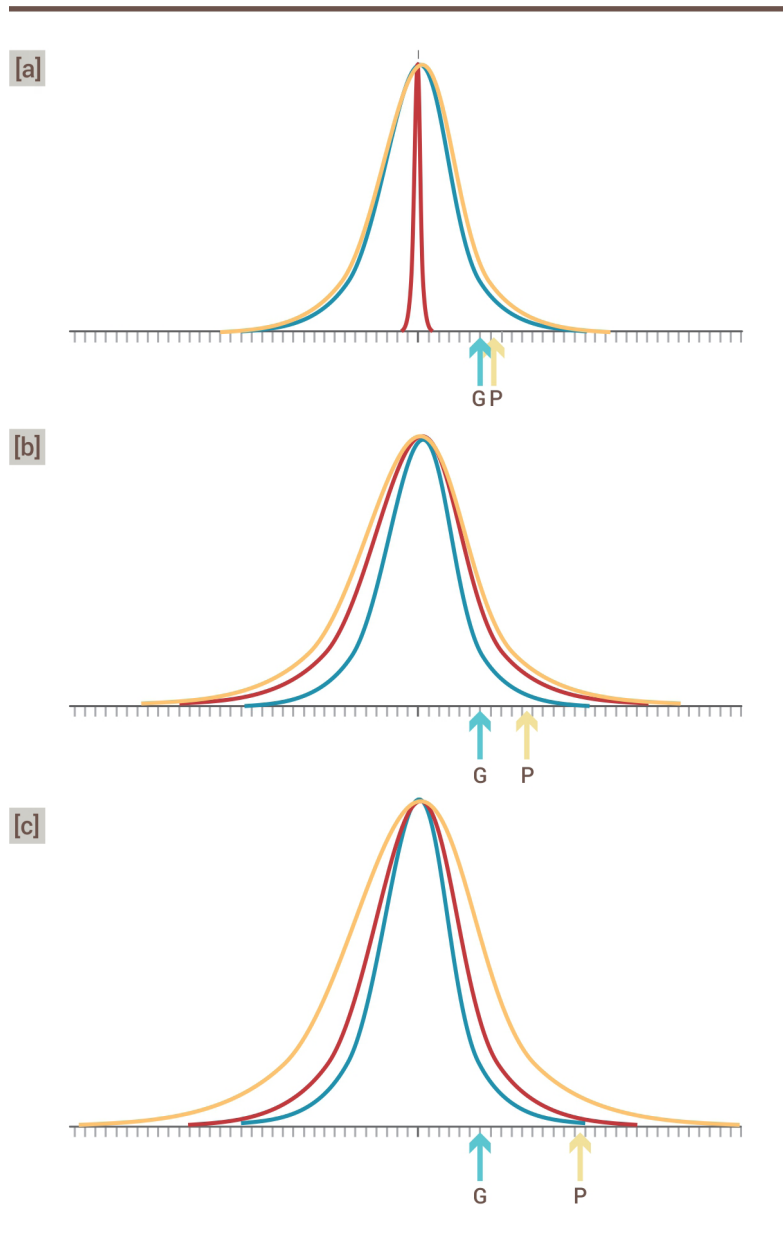



Figura 3.1. Distribución de valores fenotípicos (P, trazo amarillo), genotípicos (G, trazo azul) y ambientales (E, trazo rojo) y localización hipotética del fenotipo de un individuo y sus componentes bajo distintas variancias de la distribución de los efectos ambientales. [a] Variancia de los desvíos ambientales cercana a cero; [b] Variancia de los desvíos ambientales mayor que en [a]; [c] Variancia de los desvíos ambientales mayor que en [b].



Cuanto más próxima a cero sean la media y la variancia de la distribución de los desvíos ambientales, menor será la diferencia entre las distribuciones de P y de G. Mayor variancia de los desvíos ambientales provoca mayor apartamiento del valor fenotípico respecto de G. En un ambiente determinado, la técnica experimental al efectuar múltiples mediciones u observaciones en cada genotipo hace que la distribución del error tenga menor variancia y que el error promedio tienda a cero. Tal caso equivale a decir que la medida de P tenderá a acercarse cada vez más a G, o que

$$\varepsilon (P) = G.$$

Al modelo genético básico [3.1] puede incorporársele un término adicional que responde a las diferencias genéticas de comportamiento de los individuos cuando se exponen a distintas condiciones ambientales. Este término se denomina interacción genotipo x ambiente (IGE). En tal caso la expresión correspondiente es:

$$P=G+E+IGE \quad [3.2]$$

La Genética Cuantitativa se interesa por conocer de qué manera estos efectos inciden sobre los parámetros (media, variancia) de una población genética.

Los avances en biología molecular permiten desafiar los supuestos de los modelos en los que se fundamenta la Genética Cuantitativa, pero no en el sentido de reemplazarlos, sino por el contrario analizar su eficiencia y robustez en un entorno que permite enriquecerlos. Walsh (2002) denomina enfoque clásico de la Genética Cuantitativa al presentado, el cual resultó exitoso, especialmente en sus aplicaciones en el mejoramiento animal o vegetal, y designa como enfoque neoclásico a aquel que incorpora al modelo clásico un componente de efectos fijos atribuibles a la asociación entre marcadores moleculares y loci de caracteres cuantitativos (QTLs). De modo tal, en el modelo neoclásico el comportamiento observado (P) de un genotipo (G) puede explicarse de la siguiente manera:

$$P=G_m+G+E+IGE \quad [3.3]$$

donde G_m indica el componente del valor genotípico atribuible a los efectos de los QTLs asociados a marcadores moleculares conocidos. El valor práctico del enfoque neoclásico está directamente relacionado con que el componente de la variancia fenotípica que depende de G_m sea relativamente más importante que la suma de los restantes términos del modelo, de naturaleza aleatoria.

Regresando al modelo más sencillo [3.1], el valor genotípico (G) responde al conjunto de efectos y modos de acción de cada alelo en todos los loci que controlan un carácter cuantitativo. Estos valores expresados en el modelo

dependen exclusivamente de las propiedades de los genes que componen el genotipo.

Para ejemplificar la aplicación del modelo consideraremos en primera instancia el caso de un locus L con dos alelos (L_1 , L_2) y asumiremos que idealmente hemos podido conocer los tres valores genotípicos verdaderos resultantes de dos padres homocigotas contrastantes (L_1L_1 y L_2L_2) y de su cruzamiento (L_1L_2), luego de controlar el efecto ambiental y el error experimental, es decir, que $E=0$. Asumamos que esos valores genotípicos fueran 100 y 50 para los homocigotas L_1L_1 y L_2L_2 , respectivamente y 87 para el heterocigota L_1L_2 .

Podemos calcular el valor promedio entre los dos homocigotas (m) y expresar los valores genotípicos de L_1L_1 , L_1L_2 y L_2L_2 no ya como valores absolutos (100, 87, 50), sino como desvíos de los valores genotípicos observados respecto de (m). Así los valores genotípicos resultan ser 25 y -25 para los homocigotas L_1L_1 y L_2L_2 , respectivamente y 12 para el heterocigota L_1L_2 . Los valores genotípicos así expresados reciben una denominación especial. El desvío entre el valor genotípico observado de cada homocigota y su promedio se denomina valor aditivo (a , $-a$) o contribución de los homocigotas y el desvío entre el valor genotípico observado para el heterocigota y el promedio de los dos homocigotas, desvío de dominancia (d) (Figura 3.2).

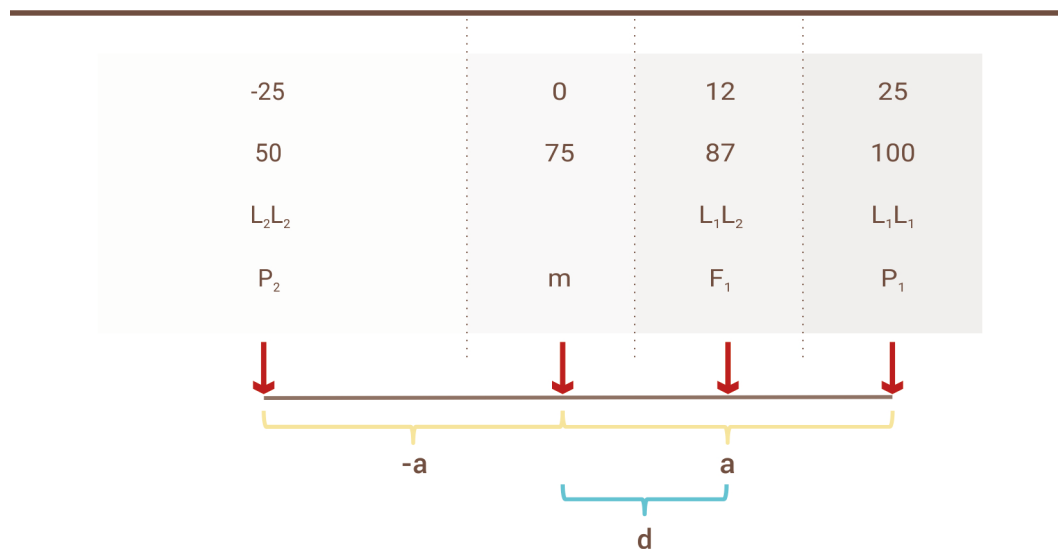


Figura 3.2. Representación gráfica, según un modelo de aditividad y dominancia, de los valores genotípicos para un locus con dos alelos, de los progenitores homocigotas para el alelo favorable (P_1) y el alelo desfavorable (P_2) y su cruzamiento (F_1). Los valores genotípicos de los progenitores son 50 y 100, el de su F_1 es 87. Los valores aditivos de los progenitores P_1 y P_2 medidos como desvío del promedio de los progenitores ($m=75$) son $a=25$ y $-a=-25$. El desvío de dominancia ($d=12$) se calcula como desvío del valor genotípico de la F_1 respecto de m .

El modelo [3.1] permite describir el valor genotípico observado a partir de las siguientes expresiones:

Para L_1L_1 :

$$G_{11} = m + a_L = 75 + 25 = 100$$

Para L_1L_2 :

$$G_{12} = m + d_L = 75 + 12 = 87$$

Para L_2L_2 :

$$G_{22} = m - a_L = 75 - 25 = 50$$

Conocidos los valores de a y d , de acuerdo a su magnitud relativa pueden darse tres situaciones. Si $|a| > |d|$ o bien $|d|/|a| < 1$, la **dominancia es parcial**. Si $|a| = |d|$ o bien $|d|/|a| = 1$, la dominancia es completa. Si $|a| < |d|$ o bien $|d|/|a| > 1$ hay sobredominancia. Estas relaciones se denominan grado de dominancia.

Continuando con el supuesto de $E=0$ y extendiendo el modelo [3.1] al caso de un carácter cuantitativo controlado por n loci (L, M, \dots, Z) se llega a las

$$\sum_{i=1}^n m' + [a] + [d] \quad [3.4]$$

siguientes expresiones:

donde

$$m' = \sum_{i=1}^n m ;$$

$$[a] = \sum_{i=1}^n a ;$$

$$[d] = \sum_{i=1}^n d$$

Para $L_1L_1M_1M_1\dots Z_1Z_1$:

$$G = m_L + m_M + m_N + \dots + m_Z + a_L + a_M \dots + a_Z$$

Para $L_1L_2M_1M_2\dots Z_1Z_2$:

$$G = m_L + m_M + m_N + \dots + m_Z + d_L + d_M \dots + d_Z$$

Para $L_2L_2M_2M_2\dots Z_2Z_2$:

$$G = m_L + m_M + m_N + \dots + m_Z - a_L - a_M \dots - a_Z$$

Para $L_1L_2M_1M_2\dots Z_1Z_1$:

$$G = m_L + m_M + m_N + d_L + d_M \dots + a_z, \text{ etc.}$$

Estas expresiones basadas en el modelo [3.4] (Figura 3.3) considerando varios loci involucrados en el carácter serán correctas siempre que no exista epistasis, es decir, que el valor aditivo o el desvío de dominancia en un locus no es afectado por el genotipo de otro/s loci.

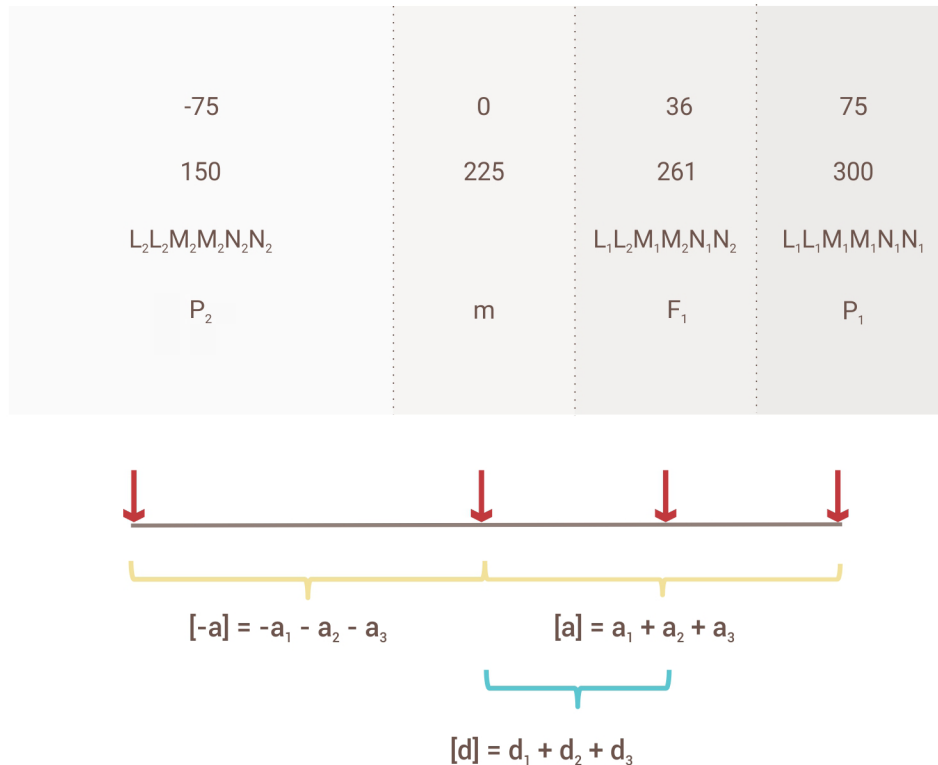


Figura 3.3.

Representación gráfica según un modelo de aditividad y dominancia de los valores genotípicos aplicado a tres loci con dos alelos de los progenitores homocigotas para los alelos favorables (P_1) y los alelos desfavorables (P_2) y su cruzamiento (F_1). Los valores genotípicos de los progenitores son 150 y 300, el de su F_1 es 261. Los valores aditivos de los progenitores P_1 y P_2 medidos como desvío del promedio de los progenitores ($m = 225$) son $[a] = 25 + 25 + 25$ y $[-a] = -25 - 25 - 25$. El desvío de dominancia ($[d] = 12 + 12 + 12$) se calcula como desvío del valor genotípico de la F_1 respecto de m .

Para describir los valores genotípicos con este modelo genético cuando se consideran los efectos de más de un locus es necesario tener en cuenta cómo se distribuyen los alelos entre los progenitores. En efecto, existen tres clases de distribución posible de los alelos favorables entre los padres. Cuando todos los loci homocigotas que llevan los alelos favorables están en uno de los padres y todos los desfavorables en el otro decimos que los genes están distribuidos en asociación o acoplamiento. Cuando la mitad de los loci homocigotas para los alelos favorables están en uno de los padres

y la otra mitad en el otro, los genes están distribuidos en repulsión o dispersión completa. La tercera alternativa ocurre cuando un número arbitrario n de loci con alelos favorables y n' desfavorables están en uno de los padres y viceversa en el otro padre y se denomina dispersión parcial o acoplamiento parcial (Tabla 3.1).

Padre	Locus						Condición	Coeficiente r_a	Valor aditivo
	A	B	C	D	E	F			
P_1	++	++	++	++	++	++	Acoplamiento	$(ka-2k'a)/ka=$	12
P_2	--	--	--	--	--	--	$k=6; k'=0$	$[(6)(2)-(2)(0)(2)]/(6)(2) = 1$	-12
P_1	--	--	--	++	++	++	Dispersión completa (50%y50%)	$(ka-2k'a)/ka=$	0
P_2	++	++	++	--	--	--	$k=6; k'=3$	$[(6)(2)-(2)(3)(2)]/(6)(2)= 0$	0
P_1	++	--	++	++	++	++	Dispersión parcial	$(ka-2k'a)/ka=$	8
P_2	--	++	--	--	--	--	$k=6; k'=2$	$[(6)(2)-(2)(2)(2)]/(6)(2) = 0,33$	-8

Tabla 3.1.

Fases de acoplamiento (asociación) y repulsión (dispersión) de genes entre dos progenitores, P_1 y P_2 , para $k=6$ loci (A, B, C, D, E y F). k' es el número de loci en repulsión. El número de loci en acoplamiento es $k-k'$. Se asume que $a=2$ para cualquiera de los loci.

Cuando todos los loci que controlan un carácter presentan desvíos de dominancia del mismo signo, se dice que la dominancia es unidireccional. Si por el contrario algunos de los desvíos de dominancia entre loci tienen signo opuesto, la dominancia es bidireccional. En este último caso eventualmente podría ocurrir que el efecto neto o sumatoria de los desvíos de dominancia a través de loci resulte en que $[d]=0$. El ejemplo representado en la Figura 3.3 considera un caso especial, en el que los genes están distribuidos en acoplamiento entre los padres y que la dominancia es unidireccional.

El componente $[a]$ del valor genotípico de los progenitores depende de la distribución de los genes, ya que toda vez que exista distribución en repulsión completa o parcial habrá que considerar el signo de la contribución de los loci tomados individualmente.

El coeficiente de asociación-dispersión (r_a) facilita el cálculo del efecto aditivo total para los progenitores de genotipo contrastante en cada uno de los

loci de interés, especialmente cuando son muchos los loci para considerar. Este se define como:

$$r_a = (ak - 2ak')/ak,$$

donde

k es el número total de loci que controla el carácter y

k' es el número de loci en repulsión.

En el caso de que el número de loci en repulsión sea 0, es decir, la condición de acoplamiento o asociación completa, el coeficiente de asociación-dispersión será igual a 1, y bajo la situación de dispersión completa, el coeficiente de asociación dispersión será igual a 0. Cuando k' es mayor de cero, es decir, bajo dispersión parcial, el coeficiente de asociación-dispersión toma un valor intermedio entre 0 y 1 (Tabla 3.1).

Empleando el coeficiente de asociación-dispersión, el efecto aditivo a través de loci $[a]$ puede calcularse como:

$$[a] = r_a k(a) \text{ y } [a] = r_a k(-a) \text{ para los padres } P_1 \text{ y } P_2, \text{ respectivamente.}$$

El desvío de dominancia a través de loci $[d]$ será:

$$[d] = \sum d_i$$

Nótese que el tamaño del desvío de dominancia no depende del coeficiente de asociación-dispersión.

En el caso de la interacción entre loci, esta varía dependiendo de la distribución de los genes favorables y desfavorables entre los padres (Kearsey y Pooni, 1996).

Cuando se consideran varios loci, no es posible calcular el grado de dominancia, puesto que $[a]$ y $[d]$ son el resultado neto o sumatoria de los efectos a y d a través de los loci que controlan el carácter. Por tal motivo la relación $[d]/[a]$ se denomina grado medio de dominancia.

3.2. Interacción epistática

Cuando dos o más loci interactúan entre sí, el valor genotípico de un individuo para un locus depende del genotipo en otro locus. Esta definición, conceptualmente biológica, se denomina epistasis fisiológica (Cheveraud y Routman, 1995).

Dados dos loci, L y M, cada uno con dos alelos, podemos construir una tabla de doble entrada con los valores genotípicos de los individuos considerando ambos loci conjuntamente (Tabla 3.2).

Genotipo	L_1L_1	L_1L_2	L_2L_2	Medias genotípicas para los genotipos ij del locus M a través del locus L = $G_{ij..}$
M_1M_1	$G_{M_1M_1L_1L_1}$	$G_{M_1M_1L_1L_2}$	$G_{M_1M_1L_2L_2}$	$G_{M_1M_1..}$
M_1M_2	$G_{M_1M_2L_1L_1}$	$G_{M_1M_2L_1L_2}$	$G_{M_1M_2L_2L_2}$	$G_{M_1M_2..}$
M_2M_2	$G_{M_2M_2L_1L_1}$	$G_{M_2M_2L_1L_2}$	$G_{M_2M_2L_2L_2}$	$G_{M_2M_2..}$
Medias genotípicas para los genotipos kl del locus L a través del locus M = $G_{..kl}$	$G_{..L_1L_1}$	$G_{..L_1L_2}$	$G_{..L_2L_2}$	Media General = $G_{....}$
Genotipo	L_1L_1	L_1L_2	L_2L_2	Medias genotípicas para los genotipos ij del locus M a través del locus L = $G_{ij..}$
M_1M_1	45	40	30	38,3
	$G_{M_1M_1L_1L_1}$	$G_{M_1M_1L_1L_2}$	$G_{M_1M_1L_2L_2}$	$G_{M_1M_1..}$
M_1M_2	50	35	25	36,7
	$G_{M_1M_2L_1L_1}$	$G_{M_1M_2L_1L_2}$	$G_{M_1M_2L_2L_2}$	$G_{M_1M_2..}$
M_2M_2	30	30	30	30
	$G_{M_2M_2L_1L_1}$	$G_{M_2M_2L_1L_2}$	$G_{M_2M_2L_2L_2}$	$G_{M_2M_2..}$
Medias genotípicas para los genotipos kl del locus L a través del locus M = $G_{..kl}$	41,7	35	28,3	Media General = 35 = $G_{....}$
	$G_{..L_1L_1}$	$G_{..L_1L_2}$	$G_{..L_2L_2}$	

Tabla 3.2.

Valores genotípicos para dos loci, cada uno con dos alelos, considerados conjuntamente.

El valor genotípico G_{ijkl} corresponde a cada una de las 3^2 celdas de la Tabla 3.2 y $G_{....}$ es la media general.

Pueden calcularse los promedios de los valores genotípicos posibles para el primer locus a través del segundo ($G_{ij..}$).

Para el caso del genotipo M_1M_1 :

$$G_{ii..} = (1/3) [G_{M_1M_1L_1L_1} + G_{M_1M_1L_1L_2} + G_{M_1M_1L_2L_2}]$$

Para el caso del genotipo M_1M_2 :

$$G_{ij..} = (1/3) [G_{M_1M_2L_1L_1} + G_{M_1M_2L_1L_2} + G_{M_1M_2L_2L_2}]$$

Para el caso del genotipo M_2M_2 :

$$G_{jj..} = (1/3) [G_{M_2M_2L_1L_1} + G_{M_2M_2L_1L_2} + G_{M_2M_2L_2L_2}]$$

Análogamente para el segundo locus a través del primero ($G_{..kl}$).

Para el caso del genotipo L_1L_1 :

$$G_{..kk} = (1/3) [G_{M_1M_1L_1L_1} + G_{M_1M_2L_1L_1} + G_{M_2M_2L_1L_1}]$$

Para el caso del genotipo L_1L_2 :

$$G_{..kl} = (1/3) [G_{M_1M_1L_1L_2} + G_{M_1M_2L_1L_2} + G_{M_2M_2L_1L_2}]$$

Para el caso del genotipo L_2L_2 :

$$G_{..ll} = (1/3) [G_{M_1M_1L_2L_2} + G_{M_1M_2L_2L_2} + G_{M_2M_2L_2L_2}]$$

El valor genotípico no epistático será:

$${}^{NE}G_{ijkl} = G_{ij..} + G_{..kl} - G_{...}$$

El valor genotípico epistático se estima como un desvío respecto del valor genotípico y será:

$${}^EG_{ijkl} = G_{ijkl} - {}^{NE}G_{ijkl}$$

A manera de ejemplo consideremos los valores indicados en la Tabla 3.2:

El valor genotípico no epistático para $G_{M_1M_2L_1L_2}$ será

$${}^{NE}G_{ijkl} = G_{ij..} + G_{..kl} - G_{...}$$

$${}^{NE}G_{M_1M_2L_1L_2} = 36,7 + 35 - 35 = 36,7$$

y su valor genotípico epistático será:

$${}^EG_{ijkl} = G_{ijkl} - {}^{NE}G_{ijkl}$$

$${}^EG_{M_1M_2L_1L_2} = 35 - 36,7 = -1,7$$

Cuando $E_{ijkl} \neq 0$ implica que existe epistasis fisiológica. La epistasis fisiológica es propiedad de los genotipos, no importan las características de la población de la que forman parte, a diferencia de la epistasis estadística (Cheveraud y Routman, 1995).

[a]				[c]			
	bb	bB	BB		bb	bB	BB
aa	8	10	12	aa	8	10	12
aA	9	11	13	aA	9	12	14
AA	10	12	14	AA	10	12	14

[b]				[d]			
	bb	bB	BB		bb	bB	BB
aa	8	10	12	aa	8	12	12
aA	9	13	17	aA	9	14	14
AA	10	16	22	AA	10	14	22

Tabla 3.3. Ejemplos de situaciones de ausencia de epistasis [a], y presencia de epistasis aditiva x aditiva [b], aditiva x dominante [c] y dominante x dominante [d].

En la Tabla 3.3 [a] podemos comprobar fácilmente la ausencia de interacción epistática porque las sumas de cualquier diagonal izquierda y diagonal derecha resultan iguales (por ejemplo: $8+11+14=10+11+12$; $8+14=12+10$; $8+11=10+9$; $11+14=12+13$). La epistasis de tipo aditivo x aditivo se comprueba observando que la suma de las diagonales que involucran a los cuatro homocigotas son distintas: ($8+22 \neq 10+12$) (Tabla 3.3 [b]). En el caso de la epistasis dominante x aditiva, se debe considerar un homocigota y un heterocigota para un locus con un homocigota y heterocigota para el otro locus ($9+14 \neq 10+14$; $8+14 \neq 9+12$) (Tabla 3.3 [c]). La epistasis dominante x dominante se comprueba cuando difiere la diagonal que involucra al heterocigota para ambos loci y al homocigota para ambos con la diagonal que suma a los heterocigotas para cada locus ($14+22 \neq 14+14$; $14+8 \neq 12+9$) (Tabla 3.3 [d]).

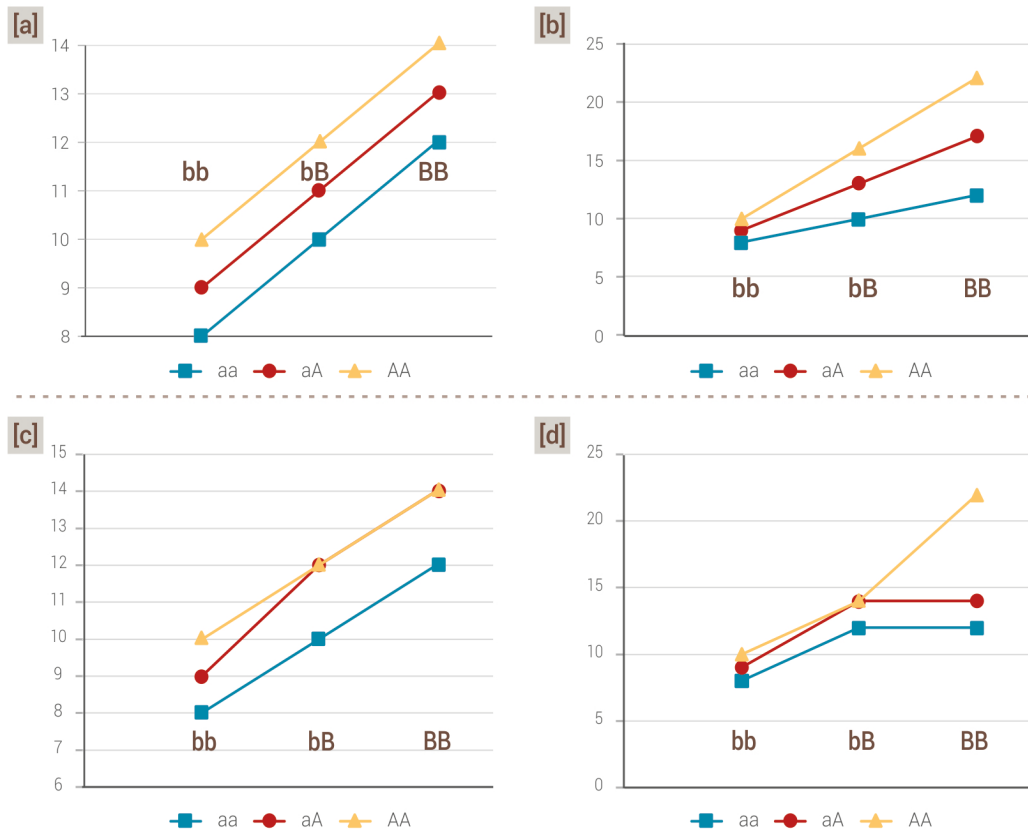


Figura 3.4. Representación gráfica de los casos de epistasis presentados en la Tabla 3.3. Las figuras [a] a [d] se corresponden respectivamente con las de la Tabla 3.3.

Lo indicado en la Tabla 3.3 se representa gráficamente en la Figura 3.4.

3.3. Epistasis estadística

La epistasis estadística responde a las frecuencias génicas poblaciones y existe cuando la interacción entre loci genera un componente de variancia genotípica (Cheveraud y Routman, 1995). El tema se presenta con mayor detalle en el Capítulo 4.

3.4. Medias generacionales

Nos referimos en esta sección a las generaciones que se obtienen por autofecundación o retrocruzamientos siguientes al cruzamiento entre dos progenitores homocigotas que son polimórficos o difieren en la constitución genética en uno o más loci que controlan un carácter. Resulta posible

deducir sus frecuencias genotípicas mediante la aplicación de las leyes mendelianas, así como también calcular la media de cada generación. Esta última será la media de los valores genotípicos de cada uno de los individuos presentes en cada generación ponderados (multiplicados) por sus frecuencias genotípicas, aplicando la fórmula [4] del Anexo.

Dados dos progenitores homocigotas polimórficos para un locus, $P_1 = L_1L_1$ y $P_2 = L_2L_2$, su cruzamiento dará origen a la filial F_1 de constitución genotípica $F_1 = L_1L_2$.

Considerando el modelo genético [3.4], los valores genotípicos serán:

$$P_1 = m + a_L$$

$$P_2 = m - a_L$$

$$F_1 = m + d_L$$

3.4.1. Medias generacionales bajo sistemas de endocría (autofecundación y retrocruzamiento)

Por autofecundación de la F_1 (o apareamiento aleatorio entre individuos genéticamente idénticos de la F_1) se obtiene la filial F_2 , cuyas características se presentan en la Tabla 3.4.

Clase Genotípica	Frecuencia genotípica (1)	Valor Genotípico (2)	(1) X (2)
L_1L_1	1/4	$m + a_L$	$\frac{1}{4} m + \frac{1}{4} a_L$
L_1L_2	1/2	$m + d_L$	$\frac{1}{2} m + \frac{1}{2} d_L$
L_2L_2	1/4	$m - a_L$	$\frac{1}{4} m - \frac{1}{4} a_L$
Media F_2			$m + \frac{1}{2} d_L$

Tabla 3.4.

Frecuencias y valores genotípicos de individuos F_2 derivados por autofecundación de individuos F_1 del cruzamiento entre líneas homocigotas y polimórficas entre sí para el locus L y media esperada de la población F_2 .

La segregación en el locus heterocigota da lugar a tres clases genotípicas, cuya abundancia o frecuencia relativa surge del producto de las frecuencias génicas (gaméticas) correspondientes. La media de la filial F_2 puede estimarse a partir de la sumatoria de los productos de tales frecuencias y valores genotípicos a través de todas las clases genotípicas (cigóticas) posibles en la F_2 . Ambos genotipos homocigotas autofecundados generan

otros individuos homocigotas en iguales frecuencias y por lo tanto sus contribuciones a la media de la generación F_2 se compensan entre sí.

La media de las siguientes generaciones o filiales que se obtengan por autofecundaciones sucesivas a partir del mismo cruzamiento, considerando un locus con dos alelos, puede estimarse a partir de similar razonamiento o mediante la aplicación de la siguiente fórmula general:

$$M_{Fn} = m + (1/2)^{n-1} d \quad [3.5],$$

o bien

$$M_{Fn} = m + (1/2)^s d \quad [3.6],$$

donde s es el número de autofecundaciones a partir de la F_1 y n indica la filial.

Conforme aumente el número de generaciones de autofecundación y siempre que el desvío de dominancia actúe en la dirección favorable al carácter, la media de la población irá disminuyendo como resultado de la disminución de la frecuencia de los individuos heterocigotas.

Las medias de las generaciones resultantes de los retrocruzamientos de la F_1 hacia uno u otro de los progenitores pueden estimarse de manera análoga (Tabla 3.5).

Padre Recurrente	Clase Genotípica	Frecuencia genotípica (1)	Valor genotípico (2)	(1) X (2)
P_1	L_1L_1	1/2	$m + a_L$	$\frac{1}{2} m + \frac{1}{2} a_L$
	L_1L_2	1/2	$m + d_L$	$\frac{1}{2} m + \frac{1}{2} d_L$
	Media de RC_1P_1			$m + \frac{1}{2} a_L + \frac{1}{2} d_L$
P_2	L_2L_2	1/2	$m - a_L$	$\frac{1}{2} m - \frac{1}{2} a_L$
	L_1L_2	1/2	$m + d_L$	$\frac{1}{2} m + \frac{1}{2} d_L$
	Media de RC_1P_2			$m - \frac{1}{2} a_L + \frac{1}{2} d_L$

Tabla 3.5.

Medias de la generación de primer retrocruzamiento (RC_1) a los padres homocigotas para el alelo favorable (P_1) y para el alelo desfavorable (P_2) bajo un modelo de aditividad y dominancia.

Considerando un locus con dos alelos, la media de cualquier generación que resulte de un número (r) de retrocruzamientos a partir del cruzamiento F_1 entre dos progenitores homocigotas puede estimarse mediante la aplicación de la siguiente fórmula general:

$$M_{RC.P1} = m + [1 - (1/2)^r] a + (1/2)^r d \quad [3.7]$$

$$M_{RC.P2} = m - [1 - (1/2)^r] a + (1/2)^r d \quad [3.8]$$

Resulta de estas expresiones que a medida que aumenta el número de retrocruzamientos va disminuyendo la frecuencia y contribución de los individuos heterocigotas a la media poblacional y va aumentando la frecuencia y contribución de los homocigotas.

3.4.2. Medias generacionales derivadas del cruzamiento entre padres homocigotas contrastantes bajo apareamiento aleatorio

Cuando las generaciones posteriores a la F_2 no se producen por autofecundación, sino por apareamiento aleatorio, la disminución de la heterocigosidad (proporción de individuos heterocigotas) es mucho más lenta y sigue los términos de la serie de Fibonacci (f_n). El coeficiente (n) es el orden del término de la sucesión, correspondiendo $n= 2$ a la filial F_2 . La media generacional será (Mather y Jinks, 1971):

$$M_{F_n} = m + (f_n) d \quad [3.9]$$

La Tabla 3.6 presenta los términos de la serie de Fibonacci para distintos valores de n , donde n es la filial. Cada término fraccionario, excepto el primero, resulta tener como denominador al doble del denominador del término precedente y como numerador a la suma de los dos numeradores precedentes.

Filial	F_2	F_3	F_4	F_5	F_6	F_7	F_8
N	2	3	4	5	6	7	8
f_n	1/2	2/4	3/8	5/16	8/32	13/64	21/128

Tabla 3.6.

Términos de la sucesión de Fibonacci que ponderan al desvío de dominancia en n generaciones de apareamiento aleatorio posteriores al cruzamiento entre dos padres homocigotas para el alelo favorable y para el desfavorable.

A manera de ejemplo, la media de una población derivada por apareamiento aleatorio entre individuos hermanos (descendientes del cruzamiento F_1) será:

$$M_{F5} = m + (5/16) d$$

La fórmula recurrente [3.9] sirve para describir cómo el apareamiento entre individuos emparentados también constituye un sistema de endocría que conduce al desarrollo de líneas homocigotas, aunque con una tasa de aproximación a la homocigosis mucho más reducida que la que corres-

ponderaría a la autofecundación de los individuos y por esta razón no tiene aplicación práctica.

Cuando existe depresión por endogamia, la media generacional para una misma filial bajo apareamiento aleatorio será mayor que bajo el esquema de autofecundación y también lo será la proporción de individuos heterocigotas.

3.4.3. Expresión de una media generacional como función lineal de otras generaciones

Mather (1949) y Mather y Jinks (1971) presentaron los sistemas de ecuaciones que expresan la media de una generación en función de las otras, bajo un esquema de autofecundación (Tabla 3.7).

Generación	P ₁	P ₂	F ₁	F ₂	F ₃	F _n	RC ₁ P ₁	RC ₁ P ₂
P ₁	1	0	0	0	0	0	0	0
P ₂	0	1	0	0	0	0	0	0
F ₁	1/2	1/2	0	0	0	0	0	0
F ₂	1/4	1/2	1/2	0	0	0	0	0
F ₂	0	0	0	0	0	0	1/2	1/2
F ₃	3/8	3/8	1/4	0	0	0	0	0
F _n	½ [1-(½) ⁿ⁻¹]	½ [1-(½) ⁿ⁻¹]	(1/2) ⁿ⁻¹	0	0	0	0	0
RC ₁ P ₁	1/2	0	1/2	0	0	0	0	0
RC ₁ P ₂	0	1/2	1/2	0	0	0	0	0

Tabla 3.7.

Expresión de medias generacionales derivadas del cruzamiento F₁ entre dos progenitores homocigotas para el alelo favorable (P₁) y desfavorable (P₂) en función de las medias de otras generaciones. F₂, F₃, y F_n son filiales obtenidas por autofecundación de la generación precedente. RC₁P₁ es la generación del primer retrocruzamiento al padre favorable y RC₁P₂ al padre desfavorable. n corresponde al número de la filial.

Así por ejemplo, la media de la generación F₃ puede calcularse como:

$$F_3 = \frac{3}{8} P_1 + \frac{3}{8} P_2 + \frac{1}{4} F_1$$

donde P₁, P₂ y F₁ son las medias de esas generaciones.

Esas relaciones entre las medias generacionales son válidas siempre que el modelo de aditividad y dominancia resulte apropiado. Para comprobarlo se utilizan pruebas de escala (Mather, 1949). En un experimento de este tipo se evalúan, por ejemplo, las generaciones $P_1, P_2, F_1, F_2, RC_1P_1$ y RC_1P_2 . Mather definió tres funciones o efectos (A, B y C) de las medias generacionales de la siguiente manera:

$$A = 2RC_1P_1 - P_1 - F_1$$

$$B = 2RC_2P_2 - P_2 - F_1$$

$$C = 4F_2 - 2F_1 - P_1 - P_2$$

donde cada uno de los términos del miembro a la derecha de la igualdad hace referencia a las medias generacionales.

Las funciones lineales, A, B y C tienen valores esperados iguales a cero, a menos que exista epistasis.

Por ejemplo:

$$A = 2RC_1P_1 - P_1 - F_1 = 2 \left[\left(\frac{1}{2}\right)a + \left(\frac{1}{2}\right)d \right] - a - d = a + d - a - d = 0$$

Sin embargo, el error de muestreo puede causar que A, B o C sean diferentes de 0, aun cuando el modelo sea suficiente y por lo tanto correcto. Para controlar esta última circunstancia es necesario tener en cuenta las variancias correspondientes a estos valores A, B y C. Estas variancias pueden calcularse según las reglas para deducir la variancia de una función lineal (ver Anexo). De tal manera serán:

$$V_A = 4V_{RC1P1} + V_{P1} + V_{F1}$$

$$V_B = 4V_{RC1P2} + V_{P2} + V_{F1}$$

$$V_C = 12V_{F2} + 4V_{F1} + V_{P1} + V_{P2}$$

Con los valores observados de A, B y C y sus correspondientes variancias puede contrastarse la hipótesis nula de que cada uno de ellos no sea significativamente distinto de cero. Para ello se realizan pruebas t de Student donde se comparan los valores de t calculados con los valores de tabla de t al nivel de probabilidad deseado y con el número de grados de libertad correspondientes. Para cada uno de las funciones A, B, o C, los grados de libertad surgen de sumar los grados de libertad de cada una de las variancias que las determinan. Así por ejemplo, si se toman $k=150$ observaciones en sendos individuos de cada una de las generaciones, los grados de libertad (g.l.) para V_A serán:

$$g.l. = 3(k-1) = 447$$

En el caso de V_c : $g.l. = 4(k-1) = 596$

A manera de ejemplo, consideremos un estudio donde se evalúan 150 individuos de las filiales P_1 , P_2 , F_1 , F_2 , RC_1P_1 , y RC_1P_2 para el carácter rendimiento por planta. Las estimaciones de los parámetros generacionales se presentan en la Tabla 3.8 y la significancia de las funciones A, B y C en la Tabla 3.9. Como puede verse la hipótesis nula de que esas funciones son diferentes de cero no puede rechazarse.

Generación	Media Observada	Variancia observada
P_1	180	225
P_2	120	256
F_1	300	380
F_2	250	2200
RC_1P_1	220	1500
RC_1P_2	190	1550

Tabla 3.8.

Medias y variancias hipotéticas de medias generacionales a partir del cruzamiento de dos progenitores homocigotas para rendimiento por planta en una especie de herencia diploide

Efecto	Media	Variancia; desvío estándar	t de Student*
A	-40	6605; 81,27	-0.492 ^{ns}
B	-40	6836; 82,68	-0.484 ^{ns}
C	100	28401; 168,52	+0.593 ^{ns}

**tg.l.*; 0.05 = 1,960; ^{ns}: no significativo

Tabla 3.9.

Media de efectos, variancias y pruebas de escala correspondientes a las generaciones indicadas en Tabla 3.7. Se asume que $k=150$ en todas las generaciones y un nivel de probabilidad $p \leq 0,05$.

En caso de que la hipótesis nula fuera rechazada, ello sería indicio de que el modelo de aditividad y dominancia no resulta suficiente para explicar las medias generacionales observadas y que existirían interacciones epistáticas entre loci.

Otras pruebas de escala pueden emplear funciones construídas a partir de un mayor número de generaciones. Ello permite ampliar el nivel de información obtenible de este tipo de análisis. Si se agregara al análisis anterior la

filial F_3 , posibilitaría definir la función D, cuya media esperada en ausencia de interacciones epistáticas sería:

$$D = 4F_3 - 2F_2 - P_1 - P_2,$$

y su correspondiente variancia:

$$V_D = 16 V_{F_3} + 4V_{F_2} + V_{P_1} + V_{P_2}$$

En caso de encontrarse que los efectos A, B, C o D son diferentes de cero existirían ciertos tipos de interacciones epistáticas. C y D proveen una prueba de escala de la presencia de efectos de interacción dominante x dominante y aditiva x aditiva, respectivamente. Los efectos A y B están asociados a interacciones del tipo aditivo x dominante.

3.4.4. Medias generacionales y epistasis

En el caso de existir interacciones epistáticas las medias generacionales resultarán entonces de la contribución de efectos genéticos de tipo aditivo, de interacción entre alelos (dominancia) y de interacciones entre loci (epistasis) (Tabla 3.10).

Generaciones	Media	Efectos genéticos				
	M	[a]	[d]	[aa]	[ad]	[dd]
P_1	1	1	0	1	0	0
P_2	1	-1	0	-1	0	0
F_1	1	0	1	0	0	1
F_2	1	0	1/2	0	0	1/4
RC_1P_1	1	1/2	1/2	1/4	1/4	1/4
RC_2P_2	1	-1/2	1/2	1/4	-1/4	1/4

Tabla 3.10.

Expresión de medias generacionales derivadas del cruzamiento F_1 entre dos progenitores homocigotas para el alelo favorable (P_1) y desfavorable (P_2) en función de los efectos genéticos de aditividad ([a]), dominancia ([d]) y de epistasis ([aa], [ad], [dd]).

Para cada generación los coeficientes que acompañan a los efectos genéticos epistáticos, aditivo x aditivo, [aa], aditivo x dominante, [ad] y dominante x dominante, [dd], resultan de la multiplicación de los coeficientes de los efectos aditivos y de dominancia correspondientes para tal generación. Por ejemplo, los coeficientes para la epistasis [aa], [ad] y [dd] en la generación RC_2P_2 serán $\frac{1}{4}$, $-\frac{1}{4}$ y $\frac{1}{4}$ puesto que los coeficientes [a] y [d] en esa generación son $(-\frac{1}{2})$ y $(\frac{1}{2})$, respectivamente:

$$M_{RC_2P_2} = m - (\frac{1}{2}) [a] + (\frac{1}{2}) [d] + (\frac{1}{4}) [aa] - (\frac{1}{4}) [ad] + (\frac{1}{4}) [dd].$$

Al igual que lo expresado para el caso de un locus, para varios loci las medias generacionales pueden combinarse en funciones lineales cuyo valor esperado es, a su vez, una función lineal de los efectos genéticos que se deseen estimar. Siguiendo a Wricke y Weber (1986):

$$2F_1 - (P_1 + P_2) = 2 [d] - [aa] + [dd]$$

Podemos considerar otras relaciones o combinaciones de interés:

$$4F_1 - 2F_1 - (P_1 + P_2) = -2[aa] - [dd]$$

$$4RC_1P_1 - 2F_1 - 2P_1 = -[aa] + [ad] - [dd]$$

$$4RC_1P_2 - 2F_1 - 2P_2 = -[aa] - [ad] - [dd]$$

$$2(RC_1P_1 - RC_1P_2) - (P_1 - P_2) = [ad]$$

$$2(RC_1P_1 + RC_1P_2) - 2F_1 - (P_1 + P_2) = -[aa] - [dd]$$

$$2(RC_1P_1 + RC_1P_2) - 4F_2 = [aa]$$

Si estas funciones resultan significativamente distintas a 0, debe concluirse que el modelo de aditividad y dominancia es insuficiente y deberían tenerse en cuenta los efectos de epistasis.

3.4.5. Media poblacional bajo condición de equilibrio de Hardy-Weinberg

Así como se pueden calcular las medias generacionales a partir del producto entre las frecuencias y los correspondientes valores genotípicos de los individuos de cada clase genotípica que la componen, la misma información permite calcular la media de una población en equilibrio de Hardy-Weinberg. En la medida que las condiciones para tal equilibrio se mantengan a lo largo de las generaciones, la media genotípica de la población permanecerá sin cambios. Considerando un locus con dos alelos

posibles, las frecuencias génicas son las que corresponden a la condición de equilibrio de Hardy-Weinberg y los valores genotípicos siguen al modelo [4.3], (Tabla 3.11).

Clase Genotípica	Frecuencia genotípica (1)	Valor Genotípico (2)	(1) X (2)
L_2L_2	q^2	- a	$-aq^2$
L_1L_2	$2pq$	d	$2pqd$
L_1L_1	p^2	a	ap^2
Media			$a(p-q) + 2dpq$

Tabla 3.11.

Media de una población en equilibrio de Hardy-Weinberg para un carácter controlado por un locus (dos alelos) cuyas frecuencias son p (alelo favorable) y q (alelo desfavorable).

La sumatoria de los valores con los que cada clase genotípica contribuye a la población, ponderados por su frecuencia, será la media poblacional expresada como desvío de la media de los homocigotas (m):

$$M = a(p-q) + 2pqr, \quad [3.10]$$

O su equivalente expresado en función de la frecuencia del alelo favorable (p):

$$M = a[p-(1-p)] + 2p(1-p)d = a(2p-1) + 2dp(1-p) \quad [3.11]$$

El miembro de la derecha de la igualdad en [3.10] posee dos componentes. El primero corresponde a la contribución a la media poblacional de los individuos homocigotas (L_2L_2 y L_1L_1) y el segundo a la contribución de los heterocigotas (L_1L_2).

La expresión de la media observada (M_o) deberá incluir el valor genotípico promedio de los dos padres homocigotas (m). Si además asumimos la ausencia de desvíos ambientales, la media observada será:

$$M_o = m + M = m + a(p-q) + 2dpq \quad [3.12]$$

En la Figura 3.5 se presentan las medias poblacionales para tres niveles de dominancia y frecuencias del alelo favorable. Nótese que la media poblacional cuando la dominancia es completa o no existe dominancia ocurre cuando el alelo favorable está fijado en la población. Cuando exista sobredominancia, la mayor media poblacional ocurrirá a alguna frecuencia intermedia.

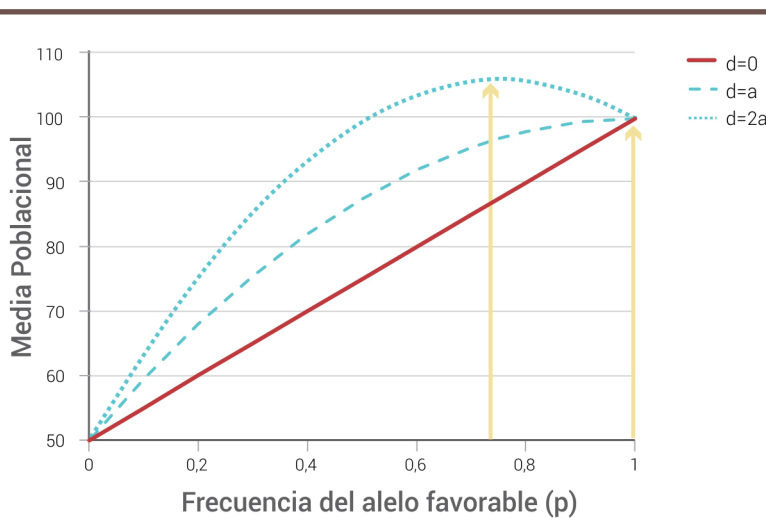


Figura 3.5. Media poblacional para diferentes frecuencias génicas, asumiendo $m=75$, $a=25$, y tres niveles de dominancia ($d=0$, $d=a$ y $d=2a$).

Las expresiones [3.10] y [3.12] pueden utilizarse también para calcular la media de la F_2 del cruzamiento entre dos progenitores homocigotas, P_1 y P_2 . Las frecuencias génicas en los loci para los cuales los padres resultan polimórficos serán $p=q=0,5$, de modo que reemplazando por esas frecuencias en las expresiones [3.10] y [3.12] llegamos a:

$$M = (\frac{1}{2})d \text{ y } M_o = m + (\frac{1}{2})d$$

En el caso de que no existan desvíos de dominancia ($d=0$) y que las frecuencias génicas sean 0,5, la media observada será simplemente el promedio de los valores genotípicos de los dos homocigotas:

$$M = M_o = m = (P_1 + P_2) / 2$$

Como estamos considerando caracteres cuantitativos, debemos tener presente que cada locus que controla el carácter hace una contribución positiva o negativa a la media poblacional. Asumiendo que el carácter estuviera condicionado por n loci y que no existieran interacciones epistáticas entre ellos, la media poblacional será:

$$M = \sum_1^n [a(p - q) + 2pqd], \quad [3.13] \text{ y}$$

$$M_o = m + \sum_1^n [a(p - q) + 2pqd], \quad [3.14] \text{ y}$$

En este caso m será el valor genotípico promedio de los dos homocigotas a través de los n loci que controlan el carácter.

Cualquiera sea el número de alelos, la media poblacional puede calcularse para un locus con múltiples alelos ($1\dots r$) mediante la sumatoria de los productos entre las frecuencias genotípicas y los valores genotípicos correspondientes a través de n loci.

$$M = \sum_{1}^n \sum_{i,j=1}^r p_i p_j g_{ij} \quad [3.15] \text{ y}$$

siendo

p_i y p_j las frecuencias alélicas en cada locus,

g_{ij} los valores genotípicos y

n el número de loci.

3.4.6. Media de poblaciones endocriadas

Cuando una población presenta endogamia, las frecuencias de genotipos homocigotas exceden, a expensas de la disminución de los heterocigotas, a las frecuencias esperadas bajo las condiciones de EHW en la misma población en ausencia de endogamia. Estos desvíos son proporcionales al coeficiente de endogamia (ver el Capítulo 5) (Tabla 3.12).

En consecuencia la media de población endogámica expresada como desvío de la media de los homocigotas (m) será:

$$M = a(p-q) + 2(1-F)dpq \quad [3.16]$$

Obsérvese que cuanto mayor sea el coeficiente de endogamia, la contribución de los heterocigotas a la media de la población será proporcionalmente menor.

La expresión de la media observada será:

$$M_o = m + M = m + a(p-q) + 2(1-F)dpq \quad [3.17]$$

Clase Genotípica	Frecuencia genotípica (1)	Valor Genotípico (2)	(1) X (2)
L_2L_2	$q^2 + pqF$	- a	$-a(q^2 + pqF)$
L_1L_2	$2pq - 2pqF = 2pq(1-F)$	d	$d[2pq(1-F)]$
L_1L_1	$p^2 + pqF$	a	$a(p^2 + pqF)$
Media			$a(p-q) + 2(1-F)d pq$

Tabla 3.12. Media de una población en equilibrio de Hardy-Weinberg con un nivel de endogamia F para un carácter controlado por un locus (dos alelos) cuyas frecuencias son p (alelo favorable) y q (alelo desfavorable).

Asumiendo que el carácter estuviera condicionado por n loci, la media poblacional será:

$$M = \sum_1^n [a(p - q) + 2pqd(1 - F)], \quad [3.18] \text{ y}$$

$$M_o = m + \sum_1^n [a(p - q) + 2pqd(1 - F)], \quad [3.19]$$

Cuando se comparan las expresiones de las medias poblacionales con y sin endocria se concluye que estas serán diferentes siempre que exista dominancia y que la diferencia entre ellas aumentará cuanto mayor sea el coeficiente de endogamia. Siempre que la dominancia sea direccional en el sentido del alelo favorable ($d > 0$) la población endogámica tendrá una media inferior a la de la misma población en condiciones de $F=0$.

Las ecuaciones [3.13] y [3.14] pueden considerarse casos especiales de [3.18] y [3.19] cuando $F=0$.

Dadas dos poblaciones, la depresión por endocria será mayor en aquella cuyas frecuencias génicas sean más próximas a 0,5 (máxima frecuencia de heterocigotas) y será menor cuanto más cercanas estén a la fijación o a la pérdida del alelo. La disminución de la media de rendimiento de la F_2 de un híbrido simple entre líneas de un híbrido de maíz, por ejemplo, es muy drástica debido a que en la F_1 la frecuencia génica es 0,5 y todos los individuos son heterocigotas.

Dada una población cuya media es:

$$M_o = m + \sum_1^n [a(p - q) + 2pqd(1 - F)] ,$$

la media de las líneas endocriadas (ML) que pudieran derivarse de ella será:

$$ML = m + \sum_1^n [a(p - q)] , \quad [3.20]$$

La media de las líneas derivables de dos poblaciones será mayor en aquella que posee mayor frecuencia de alelos favorables. En las situaciones en que un fitomejorador debe escoger entre dos o más poblaciones para ser mejoradas es importante que se preste atención al componente de la media atribuida al aporte de los individuos homocigotas. Aquella población donde esta contribución relativa sea mayor será la que permitirá la obtención de líneas endocriadas de mejor comportamiento *per se*. Es por ello que las poblaciones más seleccionadas permitirán derivar mejores líneas que las poblaciones con menor trabajo previo de mejoramiento.

3.4.6.1. Endogamia y media poblacional

A los fines de comparar la media de una misma población sin endocría (M) y con endocría (M_F) podemos utilizar la siguiente expresión:

$$M - M_F = [\sum a(p-q) + 2dpq] - [\sum a(p-q) + 2dpq(1-F)] \quad [3.21]$$

Por lo tanto

$$M_F - M = -2dpqF$$

El miembro de la derecha de la igualdad cuantifica la depresión por endocría, a condición de que exista dominancia direccional a favor del carácter. De acuerdo a esta expresión, y si existe dominancia, la depresión por endocría será máxima cuando las frecuencias génicas sean $p=q=0,5$. Por causa de tal dominancia direccional la relación entre la depresión por endocría y el coeficiente de endogamia (F) resulta ser inversamente proporcional y lineal. En el caso de existir epistasia la relación entre la media poblacional y el coeficiente de endogamia no será lineal ya que estará afectada por las interacciones entre loci heterocigotas y la declinación en la media será en función del cuadrado del coeficiente de endogamia (F^2).

3.5. Efectos medios de los alelos y efecto de sustitución de un gen

Considerando una especie diploide, el efecto medio de un gen (o de cada una de sus formas alélicas alternativas, α^+ y α^-) se define como el desvío promedio, respecto de la media poblacional de los individuos que recibieron ese alelo en cuestión de uno de sus progenitores y el otro al azar de los demás individuos de la población (Tabla 3.13).

Clases gaméticas femeninas y su frecuencia	Clases gaméticas masculinas y su frecuencia		Media de los individuos que recibieron cada clase gamética	Desvío respecto de la media poblacional o efectos medios de los alelos L^+ y L^-
	L^+, p	L^-, q		
Cigotas resultantes y su frecuencia				
$L^+, 1$	L^+L^+, p	L^+L^-, q	$ap + dq$	$q[a + (q-p)] = \alpha^+$
$L^-, 1$	L^+L^-, p	L^-L^-, q	$-aq + dp$	$-p[a + (q-p)] = \alpha^-$

Tabla 3.13. Efectos medios de los alelos favorables (L^+) y desfavorables (L^-) en una población diploide en equilibrio de Hardy-Weinberg, con frecuencias p y q respectivamente.

De la Tabla 3.13 puede deducirse que las medias de los individuos que recibieron el alelo favorable o el desfavorable de uno de sus progenitores y el restante al azar serán, respectivamente:

$$M^+ = ap + dq$$

y

$$M^- = -aq + dp$$

En consecuencia los efectos medios de los alelos favorables (α^+) y desfavorables (α^-) serán, respectivamente:

$$\alpha^+ = M^+ - M = ap + dq - [a(p-q) + 2dpq] = aq + dq(1-2p) = q[a + d(1-p-p)] = q[a + d(q-p)] \quad [3.22] \text{ y}$$

$$\alpha^- = M^- - M = -aq + dp - [a(p-q) + 2dpq] = -aq + dq(1-2p) = -p[a + d(q-p)], \quad [3.23]$$

Los efectos medios de los genes constituyen un valor asociado a sus características intrínsecas (a, d) y a la población de referencia de la que forman parte, puesto que los efectos medios dependen de las frecuencias génicas poblacionales.

El efecto de sustitución de un gen (α) se define como la diferencia entre los efectos medios de los genes (alelos) involucrados en la sustitución. Entonces:

$$\alpha = \alpha^+ - \alpha^- = q[a + d(q-p)] - \{-p[a + d(q-p)]\} = (q+p)[a + d(q-p)] = a + d(q-p) \quad [3.24].$$

Consecuentemente, los efectos medios de los genes pueden expresarse como:

$$\alpha^- = -p \alpha, \quad [3.25]$$

y

$$\alpha^+ = q \alpha, \quad [3.26]$$

El efecto de sustitución de un gen puede interpretarse como la pendiente de la regresión de los valores genotípicos en función del número de alelos favorables que contiene cada clase genotípica (Tabla 3.14 y Figura 3.6).

Clases genotípicas	Frecuencia genotípica	Valor genotípico	N.º de alelos favorables (L^+)
	f_i	Y_i	X_i
L^+L^+	p^2	a	$2p^2$
L^+L^-	$2pq$	d	pq
L^-L^-	q^2	-a	0
Medias de ($f_i \times Y_i$) y de X_i		$a(p-q) + 2dpq$	2p

Tabla 3.14.

Interpretación del efecto medio de sustitución de un gen como regresión de los valores genotípicos en función del número de alelos favorables para un locus con dos alelos.

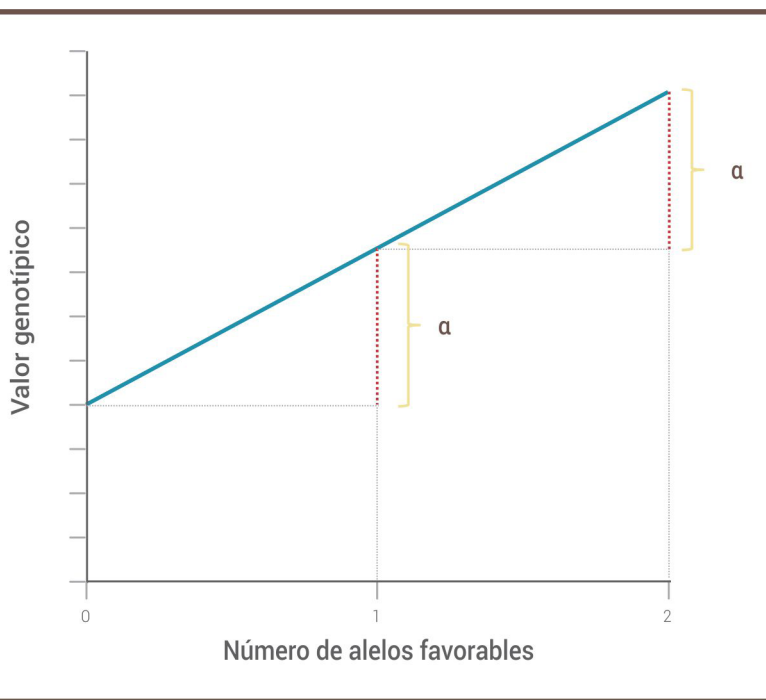


Figura 3.6. Efecto de sustitución de un gen interpretado como la pendiente de la regresión de los valores genotípicos en función del número de alelos favorables.

Estadísticamente (ver Anexo) la regresión de una variable dependiente (y) en función de otra independiente (x) se define como:

$$b_{yx} = \text{Cov}(y,x)/V_x$$

En este caso la variable independiente será el número de alelos favorables en cada clase genotípica (0, 1, 2) y la variable dependiente los respectivos valores genotípicos. Entonces en el numerador para el cálculo del coeficiente de regresión interesa la covariancia entre el número de alelos favorables y los valores genotípicos, y en el denominador la variancia para el número de alelos favorables. Es decir:

$$\begin{aligned} \text{Cov}(y,x) &= \sum_i f_i Y_i X_i - m_x m_y = p^2(m+a)2 + 2pq(1)(m+d) \\ &+ 0 - 2p[m+a(p-q) + 2pqd] \end{aligned}$$

y

$$V_x = \sum_i f_i X_i^2 - m_x^2 = p^2 2^2 + 2pq(1) + 0 - 2p^2 = 2pq$$

Reemplazando,

$$b_{yx} = a + d(q-p) = \alpha$$

La relación entre la media poblacional, los efectos medios de los genes y el efecto medio de sustitución del gen pueden visualizarse en la Figura 3.7.

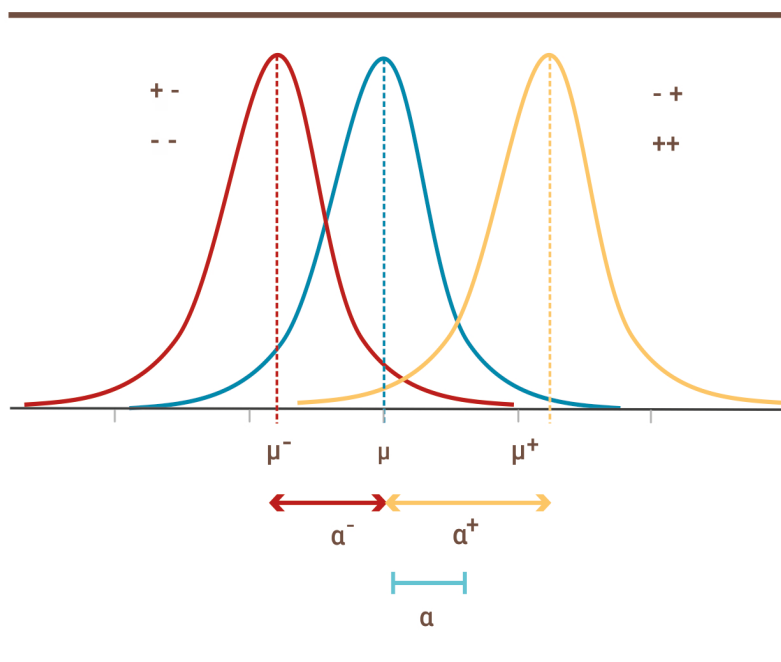


Figura 3.7. Efecto medio de dos alelos, favorable y desfavorable (α^- , α^+), y efecto de sustitución (α) para una población hipotética con una media μ . La población de la izquierda corresponde a la constituida por individuos resultantes de la fusión de gametas portadoras del alelo desfavorable con otras provenientes al azar de la población. La población de la derecha está compuesta por individuos que resultaron de la fusión de gametas con el alelo favorable con otras provenientes al azar de la población.

Cuanto mayor sea la frecuencia del alelo favorable en una población, menor será el efecto de sustitución de alelos desfavorables por favorables ya que la mayoría de los individuos de la población ya llevaban el alelo favorable. Por el contrario, cuando el alelo favorable está en baja frecuencia en una población, el efecto de sustitución de alelos desfavorables por favorables será mayor puesto que los individuos de la población que llevaban el alelo favorable eran una minoría. Esta afirmación es cierta si y solo si existe dominancia ($d \neq 0$). En ausencia de dominancia, los efectos de sustitución de un gen se mantienen constantes a través de todo el rango de frecuencias génicas (Figura 3.8).

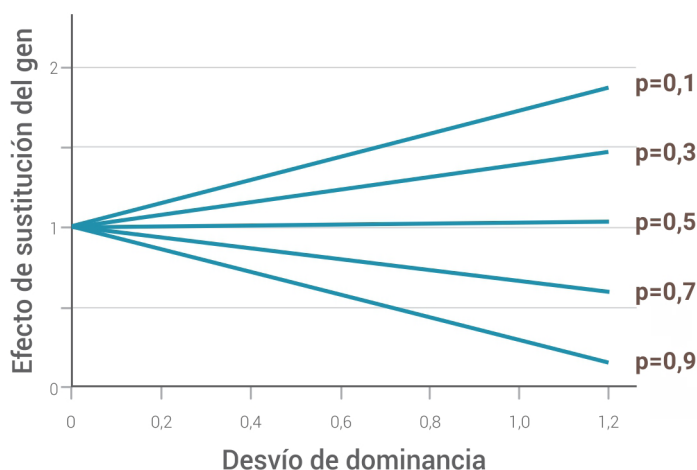


Figura 3.8. Modificación del efecto de sustitución de un gen (α) en función de la frecuencia del alelo favorable en la población (p) bajo diferentes desvíos de dominancia (d).

La media de una población en equilibrio de Hardy-Weinberg también puede expresarse en función de los efectos medios de los genes y de sustitución de un gen. Recordando que:

$$\alpha = \alpha^+ - \alpha^-$$

$$q = \alpha^+/\alpha$$

$$p = -\alpha^-/\alpha$$

y reemplazando en [3.10] la expresión final resulta ser:

$$M = -a[(\alpha^- + \alpha^+)/\alpha] - 2d[\alpha^+ \alpha^-/\alpha^2]$$

Los efectos medios de los genes y los efectos de sustitución se ven afectados por el nivel de endocria de la población. Las expresiones correspondientes para niveles arbitrarios de endogamia en una población son las siguientes:

$$\alpha_F = a + \left\{ \frac{1-F}{1+F} \right\} d(q-p) \quad [3.27]$$

$$\alpha_{F+} = q \alpha_F, \quad [3.28]$$

$$\alpha_{F-} = -p \alpha_F, \quad [3.29]$$

En una población donde $F=0$, se cumple que $^F\alpha = \alpha$. Cuando $F=1$, $^F\alpha=a$ y cuando $F=0,5$ entonces $\alpha > ^F\alpha$.

3.6. Valor de mejora de un individuo

El mérito genético de un individuo puede describirse mediante la sumatoria de los efectos medios de los genes que lo constituyen. Por definición, el valor de mejora de un individuo es el doble del desvío entre el comportamiento promedio de su progenie y la media poblacional (Falconer y Mackay, 2001). Ello se debe a que el individuo contribuye con el 50 % del genotipo de su progenie. Los términos valor reproductivo, valor aditivo y valor de mejora (*breeding value*) son conceptualmente equivalentes.

En el caso de considerar varios loci, el valor de mejora de un individuo resultará ser la sumatoria de los efectos medios de todos los genes que lo constituyen a través de los loci.

El valor genotípico de cualquier individuo puede también expresarse centrado en la media poblacional en función de los efectos de sustitución de los genes y los desvíos de dominancia (Tablas 3.15 y 3.16).

Genotipo	Genotipo de la progenie por cada alelo del individuo y frecuencia	Valor genotípico de la progenie	Desvío del valor genotípico de la progenie respecto de la media poblacional por cada alelo del individuo	Valor de mejora
L⁺L⁺	L ⁺ L ⁺ , p;		pa+qd -[a(p-q) - 2dpq]	2qa = 2α ⁺
	L ⁺ L ⁻ , q	p(a) + q(d)	= qd +aq -2dpq	
	L ⁺ L ⁺ , p;		= q[(a +d(1-2p))] = qa	
	L ⁺ L ⁻ , q	p(a) + q(d)	= qd +aq -2dpq	
L⁺L⁻	L ⁺ L ⁺ , p;		= q[(a +d(1-2p))] = qa	α(q-p) = α ⁺ + α ⁻
	L ⁺ L ⁻ , q	p(a) + q(d)	= qd +aq -2dpq	
	L ⁻ L ⁺ , p;	-q(a) + p(d)	= -p[a + d(1-2p)] = -pa	
	L ⁻ L ⁻ , q		= -p[a + d(1-2p)] = -pa	
L⁻L⁻	L ⁻ L ⁻ , q;		-aq+pd -[a(p-q) - 2dpq]	-2pa = 2α ⁻
	L ⁻ L ⁺ , p	-q(a) + p(d)	=pd - ap -2dpq=	
	L ⁻ L ⁻ , q;		= -p[a + d(1-2p)] = -pa	
	L ⁻ L ⁺ , p	-q(a) + p(d)	=pd - ap -2dpq=	
			= -p[a + d(1-2p)] = -pa	

Tabla 3.15. Valor de mejora de las clases genotípicas en un modelo simple de un locus con dos alelos.

Genotipo de los individuos	Valor genotípico	Valor genotípico en relación a la media poblacional	Valor aditivo o valor de mejora (A)	Desvío o Valor de dominancia (D)
L ⁺ L ⁺	a	$a-[a(p-q) + 2dpq] = 2q(a-dq)$	2qa	-2dq ²
L ⁺ L ⁻	d	$d-[a(p-q)-2dpq] = (q-p)a + 2pqd = \alpha(1-2p) + 2dpq$	qa - pa = $\alpha(1-2p)$	2dpq
L ⁻ L ⁻	-a	$-a-[a(p-q)-2dpq] = -2p(a+dp)$	-2pa	-2dp ²

Tabla 3.16.

Expresión de los valores genotípicos en relación con la media poblacional y su descomposición en valores de mejora y desvíos de dominancia.

Observemos que el valor genotípico de un individuo que pertenece a una población determinada puede descomponerse en la suma de dos términos, el valor aditivo o valor de mejora (A) y el valor de dominancia (D) (Tabla 3.16). Cualquier individuo, no importa su condición homocigota o heterocigota, tendrá su correspondiente valor aditivo y de dominancia, por cuanto el valor genotípico (G') se expresa en forma relativa respecto de la media poblacional:

$$G' = A + D \quad [3.30]$$

Del mismo modo, podremos definir el valor fenotípico como

$$P' = G' + E', \quad [3.31]$$

$$P' = A + D + E'$$

Como se ve, hemos podido obtener una medida de valor genotípico de cada individuo no ya en términos absolutos (a, d, -a), sino de su desvío respecto de la media poblacional. Ello resulta especialmente conveniente cuando practiquemos selección direccional, ya que necesitaremos disponer de una medida del mérito genético de cada individuo en relación con los demás que comparten la misma población. Los individuos para seleccionar serán los que posean mayor valor de mejora o valor aditivo. El valor aditivo depende de los efectos medios de los alelos que lleva el individuo, por lo que la dominancia, a través de d también influye sobre su magnitud. El valor de dominancia, en cambio, depende exclusivamente de las frecuencias génicas y los desvíos de dominancia.

En este momento podemos regresar al modelo inicialmente descrito [3.1] y concluir la relación que guarda con el modelo [3.31]:

En el modelo inicial, $P = G + E$, el valor genotípico de un individuo está expresado en función de los efectos intrínsecos del gen (a , d y $-a$) que por lo tanto no se modifican por la población a la que pertenece el individuo.

En el modelo recientemente presentado $P' = G' + E' = A + D + E'$ los valores fenotípicos, genotípicos y desvíos ambientales están centrados respecto de la media de la población a la que pertenece el individuo.

Veremos en el siguiente capítulo cómo la variabilidad entre los individuos de una población da lugar a las variancias genotípica, aditiva y de dominancia, cada una de ellas reflejando la variabilidad o dispersión de los valores genotípicos, los valores aditivos y los valores de dominancia entre los individuos de una población.

3.7. Predicción genómica del valor de mejora de un individuo

Meuwissen *et al.* (2001) propusieron un análisis de regresión de los valores fenotípicos en función de los marcadores moleculares presentes en un individuo, de acuerdo a la siguiente expresión (tomada de De los Campos *et al.*, 2013):

$$Y_i = f(x_{i1} + x_{i2} + x_{i3} + \dots x_{in}) + \epsilon_i$$

donde

Y_i es la predicción del valor de mejora del i -ésimo individuo estimado a partir de la sumatoria de los efectos de los marcadores moleculares,

$(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})$ son las estimaciones de los efectos de cada uno de los n marcadores moleculares del individuo i -ésimo, y

ϵ_i es la suma del residual no explicado por el modelo de regresión y el error experimental.

Se asume que el número n de marcadores es suficientemente elevado (se emplea una alta densidad de marcadores) y que por lo tanto la fase en la que ocurre el ligamiento entre marcadores o entre bloques de marcadores (haplotipos) y polimorfismos son consistentes a través de familias (Heffner *et al.*, 2009).

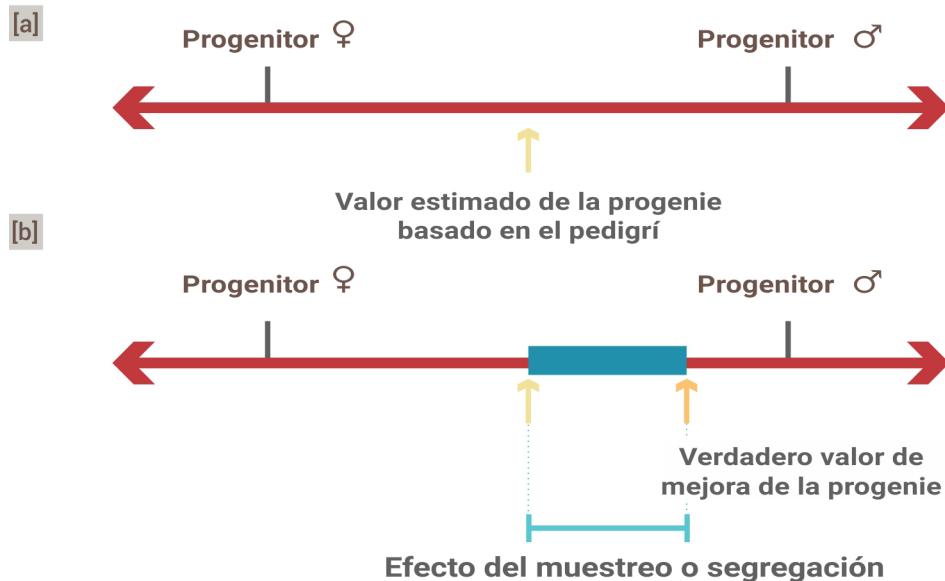


Figura 3.9. Representación del valor de mejora estimado de un individuo a partir de sus progenitores [a] y del efecto de la segregación mendeliana sobre el verdadero valor de mejora [b].

En el modelo estándar o infinitesimal la estimación del valor de mejora de una progenie se basa en el pedigrí de cada progenie y la matriz de covariancia entre parientes es la matriz de covariancia esperada (por cuanto no tiene en cuenta qué ocurre con la segregación mendeliana). En el modelo de estimación del valor de mejora basado en marcadores moleculares, la matriz de covariancia entre parientes no es la matriz esperada, sino la realizada debido a que precisamente tiene en cuenta la segregación mendeliana. Desde este punto de vista, sería factible obtener estimaciones del valor de mejora más precisos o cercanos al verdadero valor de mejora según el modelo basado en marcadores moleculares que empleando estimadores o predictores de medias familiares (Figura 3.9).

El modelo genético que expresa el valor genotípico (G) a partir de los efectos de los marcadores moleculares es el siguiente:

$$G = Q + R$$

donde

Q es la matriz de los efectos genéticos correlacionados con los marcadores moleculares a causa del desequilibrio de ligamiento y

R es la matriz de efectos genéticos aditivos que son independientes de los marcadores o que se deben a marcadores en desequilibrio de ligamiento incompleto (Dekkers, 2007).

Consecuentemente los valores fenotípicos serán:

$$P = G + E = (Q + R) + E$$

siendo E la matriz de efectos ambientales aleatorios.

Toda vez que pretendamos estimar el valor aditivo asociado a los marcadores (Q), debemos considerar que existirá un error de predicción (e). Definiremos entonces al **valor de mejora genómico verdadero** de un individuo como (Dekkers, 2007):

$$Q = Q^* + e$$

donde Q^* es el estimador del valor de mejora genómico (EVMG).

Este resulta de las estimaciones de los efectos de los marcadores alélicos provenientes del padre y de la madre de cada individuo que compone la población, sumado a través de los p marcadores considerados:

$$Q^* = \sum_p [(^m g^*) + (^p g^*)]$$

siendo

$^m g^*$ la estimación de los efectos de los marcadores alélicos del individuo recibidos de su madre y

$^p g^*$ la estimación de los efectos de los marcadores alélicos del individuo recibidos de su padre.

Volviendo al modelo sencillo, $P = G + E$, al emplear marcadores moleculares tendremos que:

$$P = G + E = (Q + R) + E = Q^* + e + R + E$$

El verdadero valor de mejora genómico (VVMG) será la sumatoria del EVMG, los errores de predicción de los efectos de los marcadores y los efectos no considerados por los marcadores moleculares:

$$VVMG = Q^* + e + R$$

Al fitomejorador le interesará conocer en qué medida el estimador del valor de mejora genómico (Q^*) es un buen estimador del verdadero valor de mejora genómico ($Q^* + e + R$) y por supuesto del verdadero valor de mejora del individuo (VVM).

Si hubiera una alta proporción de los efectos genéticos que no pueden ser captados por los marcadores, ya sea porque su número es limitado o hay falta de desequilibrio de ligamiento entre ellos y cada región del genoma que controla el carácter (QTL), R hará una contribución importante al valor genotípico (G). Si además existieran grandes errores de estimación (e) asociados a la estimación del valor de mejora genómico, esta se alejará del verdadero valor de mejora genómico (VVMG).

Dada una población podemos estimar la variancia genética explicada por el j -ésimo marcador (V_{qj}) de acuerdo a la siguiente expresión:

$$V_{qj} = r_j^2 V_{Gj}$$

donde

V_{Gj} es la contribución a la variancia genética total del j -ésimo QTL y

r_j^2 mide el desequilibrio de ligamiento entre el j -ésimo marcador y j -ésimo QTL.

Cuanto mayor sea el desequilibrio de ligamiento (mayor frecuencia de ocurrencia simultánea de marcador y de (QTL), mayor será la variancia genética explicada por el marcador.

Si se dispone de n marcadores, la variancia genética o aditiva total será la sumatoria de las contribuciones de los n marcadores individuales ($n=1\dots j$), de tal modo que:

$$V_{qj} = r_j^2 V_{Gj}$$

$$Q^2 = \sum_1^j V_{qj} = \sum_1^j (r_j^2 V_{Gj})$$

Cuanto mayor sea el desequilibrio de ligamiento, mayor el número de marcadores y mayor la variancia genética aportada individualmente por cada marcador, tanto mayor será la proporción de la variancia genética o aditiva explicada por los marcadores.

3.8. Media poblacional y epistasis

Los efectos genéticos cuando se consideran dos pares de genes, E y F, aplicando para cada locus individual el modelo [3.1] y agregando otros posibles efectos de interacción entre ambos loci (i) se presentan en la Tabla 3.17. A los valores genotípicos por cada locus se les agrega uno o más términos correspondientes al efecto de interacción epistática.

Genotipo	Locus F		
	F ₁ F ₁	F ₁ F ₂	F ₂ F ₂
Locus E E ₁ E ₁	m _{EF} + a _E + a _F + ia _E a _F	m _{EF} + a _E + d _F + ia _E d _F	m _{EF} + a _E - a _F - ia _E a _F
E ₁ E ₂	m _{EF} + d _E + a _F + id _E a _F	m _{EF} + d _E + d _F + id _E d _F	m _{EF} + d _E - a _F - id _E a _F
E ₂ E ₂	m _{EF} - a _E + a _F - ia _E a _F	m _{EF} - a _E + d _F - ia _E d _F	m _{EF} - a _E - a _F + ia _E a _F

Tabla 3.17.

Efectos aditivos y de interacciones intralocus e interloci considerando un modelo simple de dos loci y dos alelos por locus.

En una población en EHW, para un carácter controlado por dos loci con dos alelos, con interacción epistática entre ellos, la media poblacional puede calcularse a partir de la sumatoria de los valores de cada clase genotípica ponderados por su frecuencia poblacional (Tabla 3.18).

Clase Genotípica	Frecuencia genotípica (1)	Valor Genotípico (2)	(1) X (2)
E ₂ E ₂ F ₂ F ₂	q ² s ²	- a _E - a _F + a _E a _F	q ² s ² (- a _E - a _F + a _E a _F)
E ₁ E ₂ F ₂ F ₂	2pq s ²	d _E - a _F + d _E a _F	2pq s ² (d _E - a _F + d _E a _F)
E ₁ E ₁ F ₂ F ₂	p ² s ²	a _E - a _F + a _E a _F	p ² s ² (a _E - a _F + a _E a _F)
E ₂ E ₂ F ₁ F ₂	q ² 2rs	- a _E + d _F + a _E d _F	q ² 2rs(- a _E + d _F + a _E d _F)
E ₁ E ₂ F ₁ F ₂	2pq 2rs	d _E + d _F + d _E d _F	2pq 2rs (d _E + d _F + d _E d _F)
E ₁ E ₁ F ₁ F ₂	p ² 2rs	a _E + d _F + a _E d _F	p ² 2rs (a _E + d _F + a _E d _F)
E ₂ E ₂ F ₁ F ₁	q ² r ²	- a _E + a _F + a _E a _F	q ² r ² (- a _E + a _F + a _E a _F)
E ₁ E ₂ F ₁ F ₁	2pq r ²	d _E + a _F + d _E a _F	2pq r ² (d _E + a _F + d _E a _F)
E ₁ E ₁ F ₁ F ₁	p ² r ²	a _E + a _F + a _E a _F	p ² r ² (a _E + a _F + a _E a _F)

Tabla 3.18.

Contribución de diferentes clases genotípicas para un carácter controlado por dos loci con dos alelos por locus considerando un modelo de aditividad, dominancia y epistasis en una población en equilibrio de Hardy-Weinberg. Las frecuencias de los alelos favorables para los locus E y F son p y q, respectivamente, y la de los desfavorables q y s, respectivamente.

Operando algebraicamente y simplificando la sumatoria de la columna de la derecha de la Tabla 3.18, se llega a la siguiente expresión de la media poblacional:

$$M = a_E(p - q) + a_F(r - s) + 2pqd_E + 2rsd_F + (1-2pq)(1 - 2rs)a_Ea_F + 2(1- 2sr)pq-d_Ea_F + 2(1- 2pq)rsa_Ed_F + 4pqrsd_Ed_F \quad [3.32]$$

Los primeros cuatro términos del miembro de la derecha de la igualdad corresponden a las contribuciones de los efectos aditivos y de los desvíos de dominancia, mientras que los últimos cuatro son las contribuciones de los desvíos de interacción epistática a la media poblacional (epistasis aditiva x aditiva, dominante x aditiva, aditiva x dominante y dominante x dominante).

Por lo tanto la media observada será:

$$M_o = m_E + m_F + M = m_E + m_F + a_E(p - q) + a_F(r - s) + 2pqd_E + 2rsd_F + (1 - 2pq)(1 - 2rs)a_Ea_F + 2(1 - 2sr)pqd_Ea_F + 2(1 - 2pq)rsa_Ed_F + 4pqrsd_Ed_F \quad [3.33]$$

Cuando cualquiera de los loci está fijado (frecuencias $p=1$ o $r=1$) la contribución de la epistasis dominante x dominante y las que involucran aditividad y dominancia se hacen cero. En otras palabras, en poblaciones de especies autógamias, formadas por individuos homocigotas, en caso de existir epistasis, solo podría ser de tipo aditivo x aditivo. Si ambos loci tienen frecuencias alélicas intermedias (0,5), la contribución de los componentes epistáticos que involucran dominancia se hace máxima. Por el contrario, la contribución de la epistasis aditiva x aditiva es máxima cuando ambos loci están fijados y mínima cuando $p = r = 0,5$.

Si existe epistasis, la relación entre la media de una población bajo apareamiento aleatorio y la misma población con niveles arbitrarios de endogamia es la siguiente (Kempthorne, 1957):

$$^FM = M + F[d] + F^2[dd] + F^3[ddd] + F^4[dddd] + \dots$$

siendo

F el coeficiente de endocría,

[d] la suma de los efectos de dominancia a través de loci,

[dd] la suma de los efectos de epistasis dominante x dominante a través de loci,

[ddd] la suma de los efectos de epistasis dominante x dominante x dominante a través de loci,

[dddd] la suma de los efectos de epistasis de dominancia de cuarto orden a través de loci.

Consecuentemente, toda vez que exista epistasis involucrando dominancia, la disminución de la media de la población por efecto de la endocria no resultará lineal.

3.9. Medias de familias extraídas de poblacionales en EHW

Con mucha frecuencia, sea para su estudio o para su mejoramiento por selección, una población genética puede estructurarse en grupos de individuos emparentados o familias. Podemos calcular las medias familiares y luego, a partir del promedio de las medias familiares, obtener la media poblacional. Si la estructuración de la población se hace en tipos de familias no endocriadas, obtendremos el mismo resultado que considerando la media de los individuos que componen la población. Ello no se cumple cuando estructuramos la población en familias endocriadas. En tal caso el promedio de las familias resultará igual a la media de la población si su coeficiente de endogamia fuera igual al del tipo de familias endocriadas en que estructuramos la población (S_1 : $F=0,5$; S_2 : $F=0,75$, etc.).

Comprobaremos estas equivalencias, a manera de ejemplo, a partir de tres tipos de familias: de medios hermanos (MH), de hermanos completos (HC) y de familias que resultan de una generación de autofecundación (S_1).

Para formar las familias de medios hermanos tomaremos individuos y los cruzaremos (como progenitor común) con una serie de individuos de la misma población. Para el caso de un locus con dos alelos en la Tabla 3.19 se presenta la manera de obtener la media de cada uno de los tres tipos posibles de familias de medios hermanos y la media general de ellas.

Padre común ♀	Frec (1)	Gameta ♀	Gameta ♂		Valores genotípicos dentro de cada familia (2)			Frec. x Medias Familiares (1)x(2)
			q, E ₂	p, E ₁				
			Cigotas posibles			E ₂ E ₂	E ₁ E ₂	
E ₂ E ₂	q ²	(1) [*] _{E₂}	q _{(E₂E₂)}}	p _{(E₁E₂)}}	q(-a)	p(d)	-	q ² (pd-qa)
E ₁ E ₂	2pq	(1/2) _{E₂}	½ q _{(E₂E₂)}}	½ p _{(E₁E₂)}}	½ q(-a)	½ p(d)	-	2pq {½ [a(p-q) + d]}
		(1/2) _{E₁}	½ q _{(E₁E₂)}}	½ p _{(E₁E₁)}}	-	½ q(d)	½ p(a)	
E ₁ E ₁	p ²	(1) _{E₁}	q _{(E₁E₂)}}	p _{(E₁E₁)}}	-	q(d)	p(a)	p ² (pa-qd)

* Número entre paréntesis indica las frecuencias gaméticas.

Tabla 3.19.

Media genotípica de familias de medios hermanos en una población bajo equilibrio de Hardy-Weinberg para un locus con dos alelos.

La media de familias de medios hermanos será la sumatoria de las medias de cada tipo de familias ponderadas por su frecuencia. La sumatoria de la columna de la derecha de la Tabla 3.19 será:

$$M_{MH} = p^2(pa - qd) + 2pq\{½[a(p - q) + d]\} + q^2(pd - qa) =$$

$$= a(p - q) + 2pqd$$

Obsérvese que $M_{MH} = M$

Las familias de hermanos completos se forman a partir del cruzamiento de pares de plantas. Utilizando el mismo razonamiento que para las familias de medios hermanos, la media de familias de hermanos completos será la sumatoria de los términos de la columna a la derecha de Tabla 3.20.



Apareamientos posibles	Cigotas resultantes y frecuencia	Frecuencia x medias familiares	
	$E_2E_2(q^2)$	$E_2E_2; q^4$	$q^4(-a)$
$E_2E_2(q^2)$	$E_1E_2(2pq)$	$E_2E_2; (1/2)2pq^3$	$pq^3(-a + d)$
		$E_1E_2; (1/2)2pq^3$	
	$E_1E_1(p^2)$	$E_1E_2; p^2q^2$	$p^2q^2 d$
	$E_2E_2(q^2)$	$E_2E_2; (1/2)2pq^3$	$pq^3(-a + d)$
		$E_1E_2; (1/2)2pq^3$	
$E_1E_2(2pq)$	$E_1E_2(2pq)$	$E_2E_2; (1/4)4p^2q^2$	$2p^2q^2(d)$
		$E_1E_2; (1/2)4p^2q^2$	
	$E_1E_1(p^2)$	$E_1E_2; (1/2)2p^3q$	$p^3q(a + d)$
		$E_1E_1; (1/2)2p^3q$	
	$E_2E_2(q^2)$	$E_1E_2; p^2q^2$	$p^2q^2(d)$
$E_1E_1(p^2)$	$E_1E_2(2pq)$	$E_1E_2; (1/2)2p^3q$	$p^3q(a + d)$
		$E_1E_1; (1/2)2p^3q$	
	$E_1E_1(p^2)$	$E_1E_1; p^4$	$p^4 a$

Tabla 3.20. Media genotípica de familias de hermanos completos en una población bajo equilibrio de Hardy-Weinberg para un locus con dos alelos.

$$M_{HC} = q^4(-a) + pq^3(-a + d) + \dots p^4(a) = a(p - q) + 2pqd$$

Análogamente a lo ocurrido con las familias de medios hermanos, $M_{HC} = M$

Podemos proceder de modo similar para calcular las medias de cada familia que resulte de una generación de autofecundación de individuos de la población y a partir de ellas y sus respectivas frecuencias, calcular la media poblacional (Tabla 3.21).

Indiv.	Frec.	Gameta	Gameta	Gameta	Valores Ge- notípicos			Frec. x Medias Familiares
			E_2	E_1	E_2E_2	E_1E_2	E_1E_1	
E_2E_2	q^2	$(1)_{E_2}^*$	$q^2_{(E_2E_2)}$	-	-a	-	-	$-q^2(a)$
E_1E_2	$2pq$	$(1/2)_{E_2}$	$1/4 q^2_{(E_2E_2)}$	$1/4 pq_{(E_1E_2)}$	-a	d	a	$2pq[1/4(-a)+1/2(d)+1/4(a)] = pq(d)$
		$(1/2)_{E_1}$	$1/4 pq_{(E_1E_2)}$	$1/4 p^2_{(E_1E_1)}$				
E_1E_1	p^2	$(1)_{E_1}$	-	$p^2_{(E_1E_1)}$	-	-	a	$p^2(a)$

* Número entre paréntesis indica las frecuencias gaméticas.

Tabla 3.21.

Media genotípica de familias S1 (una generación de autofecundación) en una población bajo equilibrio de Hardy-Weinberg para un locus con dos alelos.

En este caso, la media de las familias obtenidas con una generación de autofecundación (M_{S_1}) será:

$$M_{S_1} = p^2(a) + 2pq[1/2(d)] + q^2(-a) = a(p - q) + pqd = M_{F=0,5} \quad [3.34]$$

La media de las familias endocriadas igualará a la media poblacional cuando $F=0,5$, el cual es el coeficiente de endogamia correspondiente a las familias S_1 [3.16].

Con frecuencia los mejoradores aplican métodos que permiten la obtención de líneas casi o completamente homocigotas ($F=1$), como por ejemplo los de duplicación de haploides, cultivo de anteras, desarrollo de líneas recombinantes, etc. Podemos proceder análogamente para calcular las medias de cada familia producto de la completa endocría de individuos de la población y a partir de ellas y sus respectivas frecuencias, calcular la media poblacional sometida a endocría completa (Tabla 3.22).

Indiv.	Frec.	Gameta	Gameta	Gameta	Valores Ge-			Frec. x Medias Familiares
			E ₂	E ₁	notípicos	E ₂ E ₂	E ₁ E ₂	
			Cigotas posibles					
E ₂ E ₂	q ²	(1) [*] _{E2}	q ² _(E2E2)	-	1(-a)	-	-	-q ² (a)
E ₁ E ₂	2pq	(1/2) _{E2}	½ q _(E2E2)	½ p _(E1E1)	½ (-a)	-	½ (a)	2pq[½ (-a)+½ (a)]= 0
		(1/2) _{E1}						
E ₁ E ₁	p ²	(1) _{E1}	-	p ² _(E1E1)	-	-	1(a)	p ² (a)

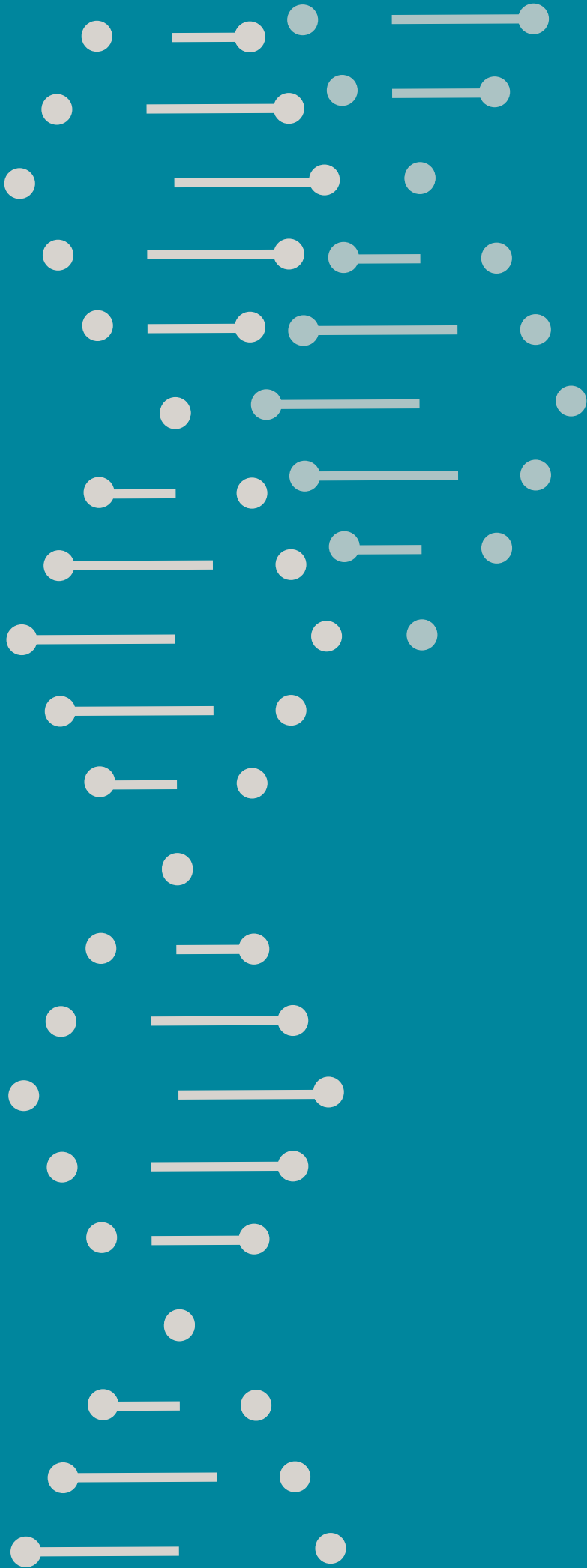
Tabla 3.22.

Media genotípica de homocigotas (F=1) en una población bajo equilibrio de Hardy-Weinberg para un locus con dos alelos.

En este caso, la media de las familias obtenidas por duplicación de haploides será:

$$M_{F=1} = p^2(a) + q^2(-a) = a(p - q) \quad [3.35]$$

La media de las familias endocriadas será igual a la media poblacional cuando F=1, conforme la expresión [3.20].



Media de cruzamientos y heterosis

4.1. Cruzamiento de poblaciones en equilibrio de Hardy-Weinberg

Al cruzar entre sí dos poblaciones, A y B, cuyas frecuencias génicas son p , q y r , s , respectivamente, las frecuencias cigóticas o genóticas resultantes del cruzamiento serán las que se presentan en Tabla 4.1. Si se suman los productos entre las frecuencias genóticas y los valores genotípicos puede comprobarse que la media del cruzamiento (M_{F_1}) es

$$M_{F_1} = M + a (pr - qs) + d (ps + qr) \quad [4.1]$$

Si reemplazamos en [4.1] a q y s por (1 - p) y (1 - r), respectivamente:

$$\begin{aligned}
 M_{F_1} &= m + a(pr - qs) + d(ps + qr) = m + a[pr - (1-r)(1-p)] + d[p(1-r) + r(1-p)] = \\
 &= m + a[pr - (1-p-r+rp)] + d[p-pr+r-rp] = m + a[-1+p+r] + d[p+r-2rp] = \\
 &= m + a[p+r-1] + d[p(1-2r)+r] \qquad \qquad \qquad [4.2]
 \end{aligned}$$

Frecuencia Gaméticas de la Población A	Frecuencias Gaméticas de la Población B	
	r	s
Frecuencias Genotípicas x Valores Genotípicos		
p	pr (a)	ps (d)
q	qr (d)	qs (-a)

Tabla 4.1. Frecuencias y valores genotípicos resultantes del cruzamiento entre dos poblaciones en equilibrio de Hardy-Weinberg con frecuencias génicas (p, q) y (r, s).

En el caso de que ambas poblaciones progenitoras tuvieran iguales frecuencias génicas (p=r; q=s), la expresión [4.1] es lógicamente equivalente a la de M [3.10]. Ello significa que la media del cruzamiento igualará a la media de cualquiera de las dos poblaciones progenitoras:

$$\begin{aligned}
 M_{F_1} &= m + a(pr - qs) + d(ps + qr) = m + a(p^2 - q^2) + d(pq + qp) \\
 &= m + a(p - q) + 2dpq \qquad \qquad \qquad [4.3]
 \end{aligned}$$

Al igual que para la media de una población, las expresiones [4.1] y [4.2] para la media de la F₁ poseen sendos términos, uno que describe la contribución de los individuos homocigotas:

$$a(pr - qs) \text{ en [4.1] y } a(p + r - 1) \text{ en [4.2]}$$

y otro de los individuos heterocigotas:

$$d(ps - qr) \text{ en [4.1] y } d(p(1 - 2r) + r) \text{ en [4.2].}$$

La media del cruzamiento interpoblacional A x B aumentará en la medida que se incrementen las frecuencias de los alelos favorables en las poblaciones progenitoras. Sin embargo cuando existe dominancia (d ≠ 0), el efecto relativo de incrementar la frecuencia del alelo favorable en una de ellas será decreciente cuanto mayor sea la frecuencia del alelo favorable de partida en la otra población progenitora (Figura 4.1).

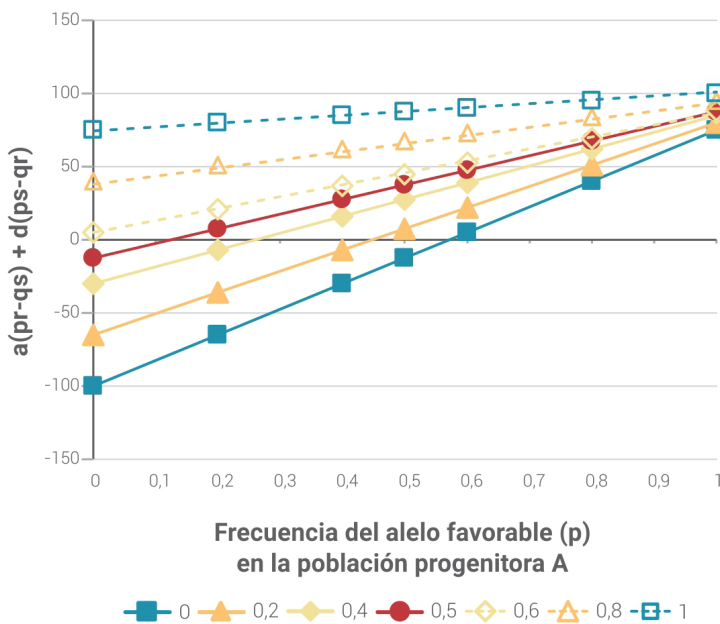


Figura 4.1. Comportamiento esperable del cruzamiento interpoblacional A x B (asumiendo $a=100$; $d=75$) para diferentes combinaciones de frecuencias génicas de los alelos favorables en la población A (p) y en la población B (r) (diferentes rectas de acuerdo a sus frecuencias génicas para r).

Esto significa que la diferencia de frecuencias alélicas entre las poblaciones progenitoras tiene incidencia importante sobre el comportamiento del cruzamiento.

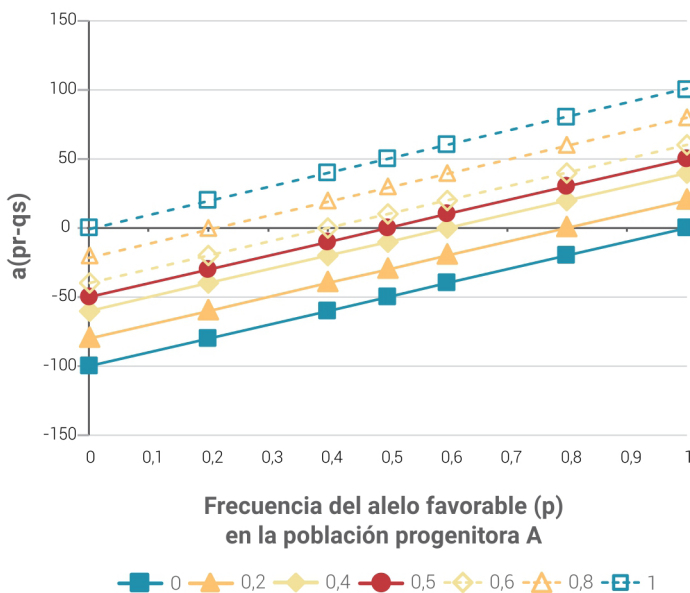


Figura 4.2. Contribución esperable de los individuos homocigotas a la media del cruzamiento interpoblacional A x B (asumiendo $a=100$; $d=75$) para diferentes combinaciones de frecuencias génicas de los alelos favorables en la población A (eje de las abscisas) y en la población B (diferentes rectas).

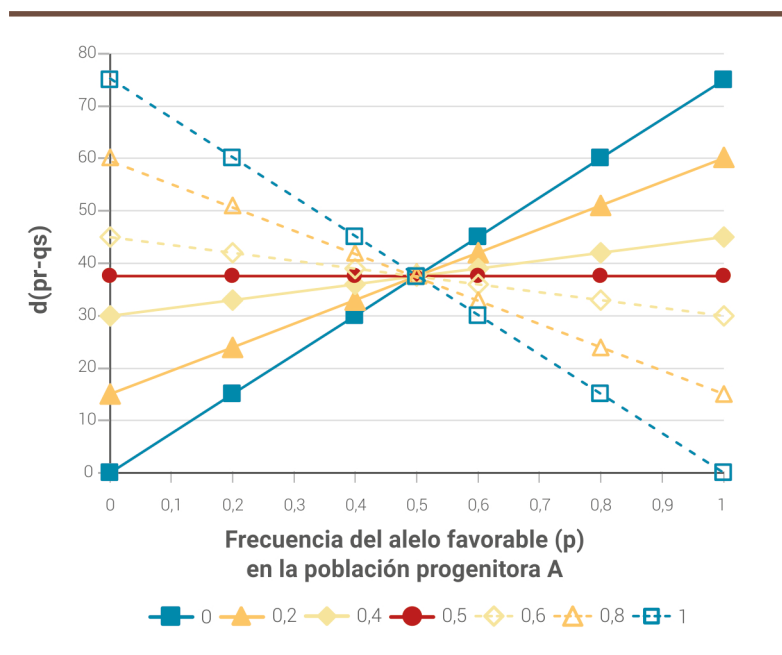



Figura 4.3. Contribución esperable de los individuos heterocigotas a la media del cruzamiento interpoblacional A x B (asumiendo $a=100$; $d=75$) para diferentes combinaciones de frecuencias génicas de los alelos favorables en la población A (eje de las abscisas) y en la población B (diferentes rectas).

En las Figuras 4.2 y 4.3 se representan las contribuciones de los individuos homocigotas y heterocigotas a la media de la F_1 interpoblacional considerando diferentes frecuencias para el alelo favorable en una de las poblaciones progenitoras (en población A: p de 0,0 a 1,0, en intervalos de 0,2); (en población B: r de 0,0 a 1,0, en intervalos de 0,2), asumiendo que $a=100$ y $d=75$.

Se observa que un incremento de la frecuencia de alelos favorables en cualquiera de las poblaciones se corresponde con un mejoramiento del comportamiento del cruzamiento F_1 . La contribución máxima de los individuos homocigotas a la media de la F_1 ocurrirá cuando ambas poblaciones hayan fijado el alelo favorable y la menor cuando este alelo esté ausente (perdido) en ambas poblaciones.

La contribución absoluta del conjunto de individuos homocigotas a la media del cruzamiento interpoblacional será mayor cuanto mayores sean las frecuencias de los alelos favorables en una o ambas poblaciones. Los efectos de los incrementos de la frecuencia alélica en una población sobre la media del cruzamiento se suman al efecto causado por los cambios de frecuencia génica en la otra población.

A mayores diferencias de frecuencias génicas entre las dos poblaciones progenitoras corresponderán mayores contribuciones de los individuos heterocigotas a la media de la F_1 interpoblacional (Figura 4.3). El aporte de los genotipos heterocigotas a la media del cruzamiento va modificándose conforme aumenta o disminuye la diferencia de frecuencias génicas para el alelo favorable entre las dos poblaciones progenitoras. Cuando las fre-



cuencias génicas en ambas poblaciones son iguales a 0,5, la contribución absoluta de los heterocigotas a la media del cruzamiento interpoblacional alcanza un punto de intersección. Si la frecuencia génica en una de las poblaciones es mayor a 0,5, la contribución absoluta de los heterocigotas a la media del cruzamiento será linealmente decreciente conforme aumente la frecuencia del alelo favorable en la otra población.

La contribución de los heterocigotas a la media del cruzamiento interpoblacional aumenta conforme la diferencia de frecuencia de los alelos favorables entre las dos poblaciones se hace cada vez más mayor (Figura 4.3.), pero el valor genotípico de la media del cruzamiento para una determinada diferencia de frecuencias para el alelo favorable entre las poblaciones progenitoras depende también de la contribución de los homocigotas, la que será mayor cuanto más cercana a 1,0 sean las frecuencias de los alelos favorables en las poblaciones. Nótese que en la Figura 4.1 para una diferencia $(p-r)=0$, la media esperada del cruzamiento F_1 será mayor cuando $p=r=1$ que cuando $p=r=0,20$.

4.2. Heterosis en cruzamientos interpoblacionales

Shull (1909, 1910) propuso el método de desarrollo de líneas puras y cruzamientos, cuya aplicación posterior permitió el desarrollo de la industria de semilla híbrida que tuvo profundos efectos en el mejoramiento de la productividad de cultivos de grano, especialmente maíz, sorgo, girasol, y especies hortícolas como por ejemplo tomate y cebolla. Años más tarde, Shull (1952) acuñó el término heterosis y su definición hace referencia al aumento en la expresión de ciertas características que se observan en los híbridos resultantes del cruzamiento, especialmente cuando los progenitores son endocriados. Falconer y Mackay (2001) definen heterosis respecto al promedio de los progenitores como la diferencia fenotípica de comportamiento del cruzamiento respecto a ese promedio y heterosis respecto al mejor progenitor como la diferencia fenotípica del híbrido respecto al comportamiento del mejor de los progenitores. La existencia de heterosis en algunas especies explica la conveniencia de organizar en grupos al germoplasma de un programa de mejoramiento genético (Sprague, 1984).

La heterosis se ha intentado explicar mediante al menos dos modelos basados en la acción génica dominante, denominados modelo de dominancia y modelo de sobredominancia (Sedcole, 1981). En el primer caso, la endocria provoca la exposición de alelos recesivos en estado homocigota que serían deletéreos para el carácter y que previamente a la endogamia estaban enmascarados en la condición heterocigota. En el segundo caso, el heterocigota tendría un valor genotípico superior a cualquiera de los pro-

genitores homocigotas, inherente a la propia condición heterocigota. La epistasis puede proveer un modelo adicional o complementario para explicar la manifestación de la heterosis. La endocría en presencia de epistasis puede modificar los efectos promedios de los alelos de un individuo. Por tal motivo, el efecto de un alelo en una línea puede ser diferente en otra o en sus cruzamientos, por expresarse en un fondo genético diferente.

Sprague y Miller (1950) y Russell y Eberhart (1975) postularon que la selección intrapoblacional por habilidad combinatoria general (sección 4.7) resultaría la más eficiente si la heterosis en maíz dependiera de la importancia de los efectos de dominancia y la selección por habilidad combinatoria específica sería más conveniente si la heterosis dependiera mayormente de los efectos de sobredominancia y epistasis. Las frecuencias génicas en el límite de selección serán diferentes en función de la importancia de esos tipos de acción génica ($p = 1$ y $p = 1 - q$), donde q es la frecuencia génica en la línea endocriada probadora). Estos autores concluyeron que la heterosis respondía fundamentalmente a efectos de aditividad y de dominancia en el rango de parcial a completa. Varios experimentos donde se evaluaron las respuestas directas e indirectas a la selección recurrente interpoblacional entre las poblaciones de maíz BS10 y BS11 (Eyhéabide y Hallauer, 1991a) y entre BSSS y BSCB1 (Keeratinijakal y Lamkey, 1993) demostraron que los incrementos en la heterosis exhibida por los cruzamientos interpoblacionales luego de varios ciclos de selección (8 y 11, respectivamente) respondían a que los cruzamientos eran cada vez más heterocigotas por la selección de diferentes alelos en cada población. Este tema será retomado en el Capítulo 9.

Puede suceder que la media del cruzamiento entre dos poblaciones (P_1 y P_2) sea superior a la de las poblaciones progenitoras. En tal situación, el comportamiento promedio del híbrido interpoblacional puede ser tal que:

$$M_{F_1} > \frac{1}{2} (M_{P_1} + M_{P_2}),$$

o bien que

$M_{F_1} > M_{P_1}$ o $M_{F_1} > M_{P_2}$, indicando en el término de la derecha de la desigualdad a la población progenitora con mayor media.

4.2.1. Heterosis respecto al promedio de los progenitores

En el primer caso señalado más arriba, la heterosis (H) respecto del promedio de los progenitores será:

$$H = M_{F_1} - \left(\frac{1}{2}\right) (M_{P_1} + M_{P_2}) \quad [4.4]$$

La media de los dos progenitores es

$$M_p = (1/2) [a(p - q) + 2dpq + a(r - s) + 2drs] = (1/2) [a(p - q + r - s) + d(pq + rs)]$$

Por lo tanto la heterosis será:

$$H = [a(pr - qs) + d(ps + qr)] - (1/2) [a(p - q + r - s) - d(pq + rs)] \quad [4.5]$$

Operando algebraicamente se llega a la expresión final:

$$H = (p - r)^2 d = (q - s)^2 d = dy^2 \quad [4.6]$$

donde

$$y^2 = (p - r)^2 = (q - s)^2$$

Obsérvese que la heterosis será mayor cuanto mayor sea la diferencia de frecuencias génicas entre las poblaciones progenitoras para el alelo favorable (o para el desfavorable) y cuanto mayor sea el desvío por dominancia. En caso de que $p=r$ y $d=0$, o una cualquiera de las dos condiciones, la heterosis será igual a cero. Asumiendo que no existe sobredominancia y que la diferencia de frecuencias génicas es diferente de cero, la heterosis será máxima si la dominancia es completa, es decir, que $d=a$.

Cuando consideramos más de un locus controlando el carácter, la heterosis será el resultado de la sumatoria de las contribuciones parciales de la heterosis en cada locus. En caso de existir dominancia bidireccional (la condición heterocigota aporta positivamente al carácter en algunos loci y en forma negativa en otros) es posible que se presente una compensación entre grupos de loci al estado heterocigota que contribuyen favorablemente al carácter con otros loci que lo hacen en forma negativa. Ello puede resultar en la existencia de niveles de heterosis no detectables experimentalmente, lo cual no puede tomarse como evidencia de la ausencia de divergencia de frecuencias génicas entre los progenitores.

Generalmente la heterosis se expresa en términos de porcentaje de superioridad de la F_1 respecto del promedio de los progenitores:

$$H = 100[M_{F_1} - (1/2)(M_{P_1} + M_{P_2})] / (1/2)(M_{P_1} + M_{P_2}) \quad [4.7]$$

4.2.2. Heterosis respecto al mejor de los progenitores

La heterosis (H) respecto del mejor de los progenitores será:

$$H = M_{F_1} - M_{P_1} \quad [4.8]$$

donde el subíndice P_1 se refiere al progenitor con la mayor media para el carácter de interés.

La heterosis respecto al mejor progenitor también puede expresarse en términos relativos como porcentaje del comportamiento del mejor progenitor:

$$H = 100 [(M_{F_1} - M_{P_1}) / M_{P_1}] \quad [4.9]$$

La media del mejor progenitor es:

$$M_{P_1} = a(p - q) + 2dpq$$

Por lo tanto la heterosis será:

$$H = a [pr - qs] + d [ps + qr] - a(p-q) - 2dpq$$

Souza y Zinsly, 1985, citados por Miranda Filho (1999), demostraron que la heterosis respecto del mejor padre es:

$${}^{MP}H = \sum (p_i - r_i)\alpha,$$

siendo

p_i y r_i las frecuencias génicas en el mejor y peor progenitor para el locus i -ésimo y

α el efecto promedio de sustitución del gen.

4.3. Heterosis panmíctica y heterosis de línea base

Lamkey y Edwards (1999) denominan heterosis panmíctica respecto del promedio de los progenitores y heterosis en la F_2 del cruzamiento entre dos poblaciones a las siguientes expresiones, respectivamente:

$${}^P H_{MP} = 4\Delta^2 d$$

y

$$H_{F_2} = 2\Delta^2 d,$$

donde

$$\Delta = \delta_1 = -\delta_2 = (p_i - p_{i'})/2 \text{ y}$$

p_i y $p_{i'}$ son las frecuencias génicas promedio de las dos poblaciones para el alelo i -ésimo.

Es decir que:

$${}^P H_{MP} = (p_i - p_{i'})^2 d$$

y

$$H_{F_2} = (p_i - p_{i'})^2 d/2$$

De estas expresiones se concluye que la heterosis se reducirá a la mitad en la F_2 respecto de la F_1 , que la heterosis en la F_1 será mayor cuanto mayor sea la diferencia de frecuencias génicas entre las poblaciones progenitoras, que la heterosis depende de la dominancia y que se modificará dependiendo de cuáles sean las características genéticas de los progenitores.

La diferencia de comportamiento de dos poblaciones parentales sin endocría (con frecuencias génicas p_1 y p_2) y las mismas dos poblaciones, pero endocriadas hasta homocigosis completa es lo que se denomina heterosis de línea base (${}^{LB}H$) y corresponde al aumento de la expresión fenotípica que se restaura en su cruzamiento interpoblacional. La heterosis respecto del promedio de los padres endocriados es la diferencia de comportamiento del híbrido respecto del comportamiento de las poblaciones parentales cuando han sido completamente endocriadas (${}^I H_{MP}$). Por lo tanto

$${}^I H_{MP} = 2p_1p_2 d + 2\Delta^2 d$$

$${}^{LB}H = 2p_1p_2 d - 2\Delta^2 d$$

$${}^I H_{MP} = {}^{LB}H + {}^P H_{MP}$$

La heterosis de línea base es la depresión por endocría promedio de las dos poblaciones progenitoras.

La heterosis panmíctica respecto al progenitor medio resulta de la acción génica de dominancia y de efectos de epistasis aditiva x aditiva entre loci independientes para los cuales los padres son divergentes. Sin embargo el comportamiento del cruzamiento F_1 depende de la dominancia y de efectos de epistasis dominante x dominante para loci independientes y para los cuales los padres son divergentes.

Para que la heterosis panmíctica respecto del promedio de los padres sea mayor que la heterosis de línea base es importante que ambos progenitores tengan frecuencias génicas en extremos opuestos uno de otro.

La Figura 4.4 ilustra claramente cuánto depende la heterosis de la diferencia de frecuencias génicas entre los progenitores, considerando un valor arbitrario de $d = 75$ y diferentes combinaciones de frecuencias génicas en una y otra población. La heterosis será máxima cuando un alelo favorable está fijado en una de las poblaciones y en la otra está perdido y será mínima cuando la frecuencia del alelo favorable es 0,5 en ambas poblaciones.

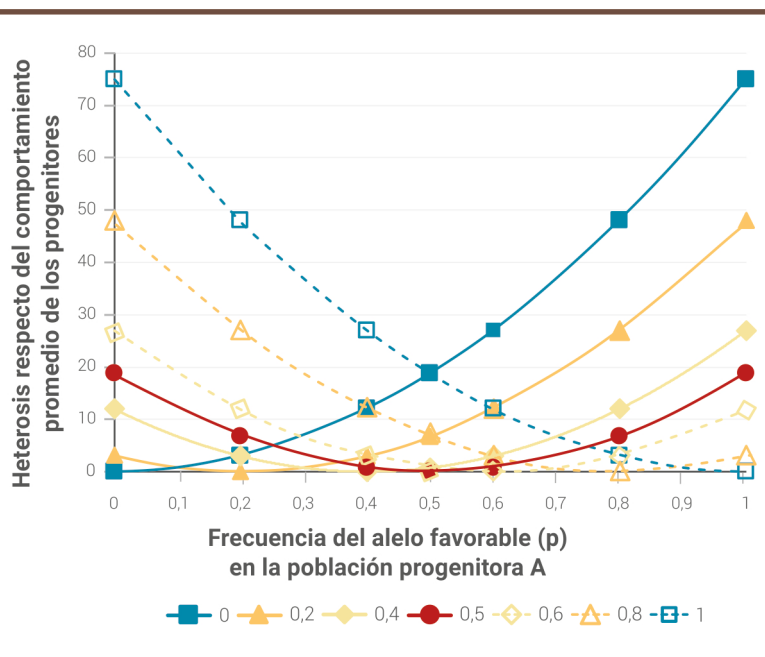


Figura 4.4. Heterosis respecto al padre medio (asumiendo $a=100$; $d=75$) esperable en el cruzamiento interpoblacional A x B para diferentes combinaciones de frecuencias génicas de los alelos favorables en la población A (eje de las abscisas) y en la población B (diferentes curvas).

Hasta el momento hemos sostenido como condiciones necesarias para la heterosis que exista dominancia y divergencia genética entre los progenitores. Esto es cierto en el caso de dos alelos por locus, sin embargo estos dos factores no son razón suficiente cuando admitimos alelos múltiples por locus. Cress, citado por Lamkey (1999) presentó el modelo correspondiente.

Si el cruzamiento F_1 entre dos poblaciones se reproduce de manera aleatoria, puede comprobarse que la población F_2 se encontrará en equilibrio de Hardy-Weinberg (EHW) (si se trata de un locus). Las frecuencias génicas en la F_2 serán el promedio de las frecuencias de ambas poblaciones: $(p + r)/2$ y $(q + s)/2$. Reemplazando en [3.16], la media de la F_2 resultará ser:

$$\begin{aligned}
 MF_2 &= a[(p + r)/2 - (q + s)/2] + 2d[(p + r)/2](q + s)/2 = \\
 &= a[(1/2)(p + r - q - s)] + 2d[(1/4)(p + r)(q + s)] = \\
 &= (1/2)a[2p + 2r - 2q - 2s] + (1/2)d(p + r)(q + s) = \\
 &= a(p + r - 1) + (1/2)d(p + r)(q + s) = a(p + r - 1) + \\
 &\quad + (1/2)d \{(p + r)[(1 - p) + (1 - r)]\} = \\
 &= a(p + r - 1) + (1/2)d \{(p + r)\{(1 - p + 1 - r)\}\} = \\
 &= a(p + r - 1) + (1/2)d \{(p + r)\{2 - p - r\}\}
 \end{aligned}$$

$$\begin{aligned}
 MP &= (1/2)[a(p - q) + 2dpq + a(r - s) + 2drs] \\
 &= (1/2)a[p - q + r - s] + 2d[pq + rs] = (1/2)a[2p - 2q + 2r] + d[pq + rs] = \\
 &= a[p + r - 1] + d[pq + rs] = a[p + r - 1] + d[p(1 - p) + r(1 - r)] =
 \end{aligned}$$

$$\begin{aligned}
&= a[p+r-1] + d[p(1-p) + r(1-r)] \\
MF_2 - MP &= \{a(p+r-1) + (\frac{1}{2})d[(p+r)(2-p-r)]\} - \{a[p+r-1] + \\
&+ d[(p(1-p) + r(1-r))]\} = \\
&= (\frac{1}{2})d[(p+r)(2-p-r)] - d[p(1-p) + r(1-r)] \\
&= d[(\frac{1}{2})(2p-p^2-pr+2r-rp-r^2)] - d[p-p^2+r-r^2] \\
&= d[p - \frac{1}{2}p^2 - \frac{1}{2}pr + r - \frac{1}{2}pr - \frac{1}{2}r^2 - p + p^2 - r + r^2] = \\
&= d[\frac{1}{2}(p^2 + r^2) - pr]
\end{aligned}$$

Tengamos en cuenta que $(p-r)^2 = (p-r)(p-r) = p^2 + r^2 - 2pr$, por lo tanto:

$$(\frac{1}{2})p^2 + (\frac{1}{2})r^2 - pr = (\frac{1}{2})(p-r)^2$$

En consecuencia:

$$MF_2 - MP = (\frac{1}{2})d(p-r)^2$$

Por lo tanto:

$$F_2H = \frac{1}{2}F_1H,$$

donde

F_2H es la heterosis respecto del progenitor medio en la filial F_2 y

F_1H es la heterosis respecto del progenitor medio en la filial F_1 .

En tanto no se modifiquen las condiciones para el EHW, la heterosis se mantendrá estable en las posteriores generaciones y su magnitud será del 50 % respecto a la verificada en la generación F_1 e idéntica a la de la F_2 .

Los fitomejoradores buscan maximizar la heterosis. Recordando que:

$$M_{F_1} = a(p+r-1) + d[p+r-2pr],$$

$$M_{P_1} = a(2p-1) + 2dp(1-p),$$

$$M_{P_2} = a(2r-1) + 2dr(1-r)$$

Asumamos para simplificar que $a = d = 1$ (Wricke y Weber, 1986):

$$M_{F_1} = a(p+r-1) + d[p(1-2r) + r] = 1 - 2qs, \quad [4.10]$$

$$M_{P_1} = a(2p-1) + d2p(1-p) = 1 - 2q^2, \quad [4.11]$$

$$M_{P_2} = a(2r-1) + d2r(1-r) = 1 - 2s^2, \quad [4.12]$$

Toda vez que $q \neq s$ la expresión [4.10] será siempre menor que las correspondientes a alguna de las medias de las poblaciones progenitoras y por lo tanto la heterosis respecto a la mejor población progenitora resultará negativa (Tabla 4.2).

Frecuencias génicas en P_1 (q) y P_2 (s)	Media de F_1 $1 - 2qs$	Media de P_1 $1-2q^2$	Media de P_2 $1-2s^2$	Promedio de los padres	H_{PM}	H_{P+}
$q=s=0,1$	0,98	0,98	0,98	0,98	0	0
$q=s=0,5$	0,50	0,50	0,50	0,50	0	0
$q=s=0,9$	-0,62	-0,62	-0,62	-0,62	0	0
$q=0,2; s=0,5$	0,80	0,92	0,50	0,71	0,09	-0,12
$q=0,5; s=0,2$	0,80	0,50	0,92	0,71	0,09	-0,12
$q=0,8; s=0,7$	-0,12	-0,28	0,02	-0,13	0,01	-0,14

Tabla 4.2.

Ejemplo de heterosis respecto del promedio de los padres (H_{PM}) y heterosis respecto al mejor padre H_{P+} para distintas combinaciones de frecuencias del alelo favorable en los progenitores de un cruzamiento interpoplacional.

Cuando consideramos varios loci, en ausencia o presencia de sobredominancia, la heterosis medida sobre el mejor progenitor puede ser positiva, aun sin tener en cuenta la eventual contribución de la epistasis (Tabla 4.2).

4.4. Heterosis y epistasis

Una tercera hipótesis para explicar el fenómeno de heterosis es la epistasis estadística (Goodnight, 1999). Aun cuando la contribución de la interacción entre determinados loci fuera importante para la endocría y heterosis, no es posible predecir esta última a partir de las características individuales de los progenitores. El fondo genético es el marco donde tales contribuciones se modifican y entonces puede suceder que loci que se comportan como neutros para la heterosis en un determinado fondo genético, pueden ser importantes en otros, y viceversa. Cuando existe epistasis estadística, la modificación del fondo genético (frecuencias génicas en otros loci) implica que los efectos medios de los genes también se modifican.

La detección de heterosis no es condición suficiente para implicar la existencia de dominancia para un carácter, sino que también es importante analizar si se verifica o no depresión por endocría. En especie autógamas, en las cuales no hay depresión por endocría, la existencia de heterosis respecto al promedio de los padres podría atribuirse a efectos de epistasis aditiva x aditiva.

4.5. Modelo de Gardner y Eberhart sobre cruza- mientos interpoblacionales

Las medias de dos poblaciones (P_i, P_j) y su cruzamiento (C_{ixj}) pueden expresarse de acuerdo a los modelos ya presentados como la suma de la media general (μ) y los componentes aditivo (A) y de dominancia (D) de cada genotipo. En el caso de las poblaciones progenitoras:

$$P_i = \mu + A_i + D_i$$

$$P_j = \mu + A_j + D_j$$

Las medias de las mismas poblaciones luego de una generación de autofecundación y de su cruzamiento interpoblacional serán (Gardner y Eberhart, 1966):

$${}^I P_i = \mu + A_i + (1/2)D_i$$

$${}^I P_j = \mu + A_j + (1/2)D_j$$

$$C_{ixj} = \mu + (1/2)(A_i + A_j) + (1/2)(D_i + D_j) + h_{ij}$$

$${}^I C_{ixj} = \mu + (1/2)(A_i + A_j) + (1/4)(D_i + D_j) + (1/2)h_{ij}$$

$${}^R C_{ixj} = \mu + (1/2)(A_i + A_j) + (1/2)(D_i + D_j) + (1/2)h_{ij}$$

donde:

${}^I C_{ixj}$ y ${}^R C_{ixj}$ son las medias del cruzamiento interpoblacional luego de autofecundar la F_1 (C_{ixj}) y de permitir su apareamiento aleatorio, respectivamente,

h_{ij} corresponde a la heterosis producto de las diferencias de frecuencias génicas entre P_i y P_j en loci que presentan dominancia.

A mayor diferencia de frecuencias génicas para loci dominantes, mayor será la contribución de la heterosis a la media del cruzamiento interpoblacional.

Cuando se realizan cruzamientos entre un grupo de más de dos poblaciones (variedades), hay suficientes grados de libertad para particionar el parámetro de heterosis h_{ij} en cuatro componentes:

$$h_{ij} = {}^P h + h_i + h_j + s_{ij}, \quad [4.13]$$

donde:

${}^P h$ es la heterosis promedio aportada por el grupo de variedades evaluadas,

h_i y h_j son, respectivamente, las heterosis varietales de la i -ésima y j -ésima población progenitora, que corresponden a la heterosis que presentan en promedio los cruzamientos de cada una de estas poblaciones. Se estima como desvío respecto de la heterosis promedio,

s_{ij} es el efecto de heterosis específica del cruzamiento entre la i -ésima y j -ésima población progenitora.

En el capítulo referido a diseños genéticos ampliaremos estos conceptos.

4.6. Medias de cruzamientos de prueba o test crosses

La media de un cruzamiento interpoblacional también puede expresarse en términos de los conceptos de efectos promedio y de sustitución de un gen, pero en cruzamientos de prueba. Suponiendo que partimos de una población (P_1) con frecuencias génicas p y q , que se cruza con una población probadora (P_2) con frecuencia génicas r y s , podríamos definir la media de los individuos generados a partir de una gameta favorable de la población (P_1) y otra tomada al azar proveniente de la población (P_2) y de manera análoga, la media de los individuos generados a partir de una gameta desfavorable de (P_1) y una gameta tomada al azar proveniente de la población (P_2), como se indica en la Tabla 4.3. En esta se representa el caso general de cruzamientos de prueba, con valores arbitrarios para p , q , r y s .

Gametas de (P_1) y su frecuencia	Frecuencia de gametas de P_2		Media	Efectos Promedios = M^+ o $M^- - M_{F_1}$
	r^+	s^-		
Frecuencias genotípicas en la F_1 x (Valor Genotípico)				
$p, 1$	$r (a)$	$s (d)$	$M^+ = ar + ds$	$\tau\alpha^+ = q[a+d(s-r)] = q[a+d(1-2r)]$
$q, 1$	$r (d)$	$s (-a)$	$M^- = -as + dr$	$\tau\alpha^- = -p[a+d(s-r)] = -p[a+d(1-2r)]$
Media F_1	$M_{F_1} = a (pr - qs) + d (ps + qr)$			
Efecto de sustitución en cruzamientos de prueba			$\tau\alpha = \tau\alpha^+ - \tau\alpha^- = a + d(s-r) = a+d(1-2r)$	

Tabla 4.3.

Efectos promedios de un gen y efecto promedio de sustitución en cruzamientos de prueba de individuos de una población en equilibrio de Hardy-Weinberg contra un probador con frecuencias génicas r (para el alelo favorable) y s (para el alelo desfavorable) aplicado al caso de un locus con dos alelos.

El valor genotípico de un individuo perteneciente al cruzamiento de la población (P_1) por otra población probadora (P_2) puede expresarse en términos de aquellos efectos promedios de sus alelos y el residual debido a dominancia (Bernardo, 2010):

$$^{P_1 \times P_2} G_{i,j} = M_{P_1 \times P_2} + \alpha_i^{P_1} + \alpha_j^{P_2} + \delta^{P_1 P_2}_{ij} \quad [4.14]$$

Los términos segundo y tercero del miembro de la derecha son los efectos promedio del alelo i de P_1 en cruzamiento de prueba con la P_2 y el efecto promedio del alelo j de la P_2 en cruzamiento de prueba con la P_1 , respectivamente. El último término es el residual debido a dominancia (asumiendo ausencia de epistasis).

Veremos que los efectos promedio y el de sustitución de un gen en cruzamientos de prueba no dependen de las frecuencias génicas de la población que se evalúa, sino de las frecuencias génicas del probador o *tester*, a las que designaremos como r y s .

En el caso de utilizar un probador híbrido simple, las frecuencias génicas en este pueden ser 0, 1, o 0,5. Los dos primeros casos corresponden a los loci en los que ambas líneas progenitoras del probador son homocigotas y homogéneas, y el tercer caso corresponde a los loci polimórficos entre las líneas progenitoras del híbrido probador. Suponiendo que hay dos alelos por locus, los dos primeros casos equivalen a utilizar una línea homocigota como probadora, sea para el alelo favorable o para el alelo desfavorable.

Cuando el probador es heterocigota para el locus de interés (Tabla 4.4), los efectos promedio de los genes en un cruzamiento de prueba dependen de las frecuencias génicas en el material probado y el efecto de sustitución del gen en cruzamiento de prueba es igual al promedio de los dos homocigotas.

Gametas de (P_1) y su frecuencia (p, q)	Gametas de (P_2) y su frecuencia		Media	Efectos Promedios = $M^+ \text{ o } M^- - M_{F_1}$
	$r^+ = 0,5$	$s^- = 0,5$		
Frecuencias genotípicas en la F_1 x (Valor Genotípico)				
p, 1	$ra = \frac{1}{2} (a)$	$sd = \frac{1}{2} d$	$M^+ = \frac{1}{2} (a + d)$	$\tau\alpha^+ = aq$
q, 1	$rd = \frac{1}{2} (d)$	$s(-a) = \frac{1}{2} (-a)$	$M^- = \frac{1}{2} (-a + d)$	$\tau\alpha^- = -ap$
Media de la F_1	$M_{F_1} = \frac{1}{2} a (p - q) + \frac{1}{2} d = \frac{1}{2} [a(p-q) + d]$			
Población x Híbridos Simple	Efecto de sustitución en cruzamientos de prueba			$\tau\alpha = \tau\alpha^+ - \tau\alpha^- = a$

Tabla 4.4.

Efectos promedios de un gen y efecto promedio de sustitución en cruzamientos de prueba de individuos de una población en equilibrio de Hardy-Weinberg contra un probador con frecuencias génicas $r=s=0,5$ (híbrido simple entre dos líneas endocriadas) para un locus con dos alelos.

En el caso de una línea probadora homocigota para el alelo favorable o el desfavorable los efectos medios y los de sustitución del gen se presentan en Tabla 4.5. Si se utiliza a la propia población como probadora, los efectos medios y de sustitución son los que se presentan en la Tabla 4.6. Si se emplea como probadora a la propia población, el efecto de sustitución de un gen en cruzamiento de prueba es igual al efecto medio de sustitución del gen en la población de referencia y los efectos medios de cada alelo en cruzamientos de prueba son iguales a los efectos medios de los alelos en la población de referencia.

Gametas de P ₁ y su frecuencia	Gametas de P ₂ y su frecuencia	Media	Efectos Promedios = Media - M _{F₁}
	r ⁺ (1)		
Frecuencias genotípicas en la F ₁ x (Valor Genotípico)			
1	(a)	M ⁺ = a	τ _{α⁺} = q (a + d)
-1	(d)	M ⁻ = d	τ _{α⁻} = -p (a - d)
Media de la F ₁	M _{F₁} = ap + dq		
Población por línea homocigota desfavorable	Efecto de sustitución de un gen en cruzamiento de prueba		τ _α = τ _{α⁺} - τ _{α⁻} = a - d

Gametas de P ₁ y su frecuencia	Gametas de P ₂ y su frecuencia	Media	Efectos Promedios = Media - M _{F₁}
	s ⁻ (1)		
Frecuencias genotípicas en la F ₁ x (Valor Genotípico)			
1	(d)	M ⁺ = d	τ _{α⁺} = aq + dq = q (d + a)
-1	(-a)	M ⁻ = -a	τ _{α⁻} = ap - dp = -p (a + d)
Media de la F ₁	M _{F₁} = -aq + dp		
Población por línea homocigota desfavorable	Efecto de sustitución de un gen en cruzamiento de prueba		τ _α = τ _{α⁺} - τ _{α⁻} = a + d

Tabla 4.5. Efectos promedios de un gen y efecto promedio de sustitución en cruzamientos de prueba de individuos de una población en equilibrio de Hardy-Weinberg contra dos probadores con frecuencias génicas r=1 y s=1 (líneas homocigotas) para un locus con dos alelos.

Gametas de P ₁ y su frecuencia	Gametas de P ₂ y su frecuencia		Media	Efectos Promedios = M ⁺ o M ⁻ - M _{F₁}
	p ⁺	q ⁻		
Frecuencias genotípicas en la F ₁ x (Valor Genotípico)				
p, 1	p (a)	q (d)	M ⁺ = ap + dq	τ _{a⁺} = q[a+d(q-p)]
q, 1	p (d)	q (-a)	M ⁻ = -aq + dp	τ _{a⁻} = -p[a+d(q-p)]
Media de la F₁	M_{F₁} = a (p - q) + 2d(pq)			
Efecto de sustitución de un gen en cruzamientos de prueba				τ_a = τ_{a⁺} - τ_{a⁻} = a + d (q-p) = α

Tabla 4.6.

Efectos promedios de un gen y efecto promedio de sustitución en cruzamientos de prueba de individuos de una población en equilibrio de Hardy-Weinberg utilizando a la misma población como probadora para un locus con dos alelos.

Resulta claro que las características del probador utilizado en términos de sus frecuencias génicas modifican los efectos medios de los alelos en cruzamientos de prueba y consecuentemente también los efectos medios de sustitución del gen. Las diferentes alternativas se presentan en la Tabla 4.7.

Efectos	Probador				
	Propia población	Población no relacionada	Híbrido simple	Línea homocigota para alelo favorable	Línea homocigota para alelo desfavorable
τ _{a⁺}	α ⁺ = q α = q[a+d(q-p)]	τ _{a⁺} = q α = q[a+d(s-r)]	τ _{a⁺} = q α = qa	τ _{a⁺} = q α = q(a-d)	τ _{a⁺} = q α = q(a+d)
τ _{a⁻}	α ⁻ = -p α = -p [a+d(q-p)]	τ _{a⁻} = -p α = -p [a+d(s-r)]	τ _{a⁻} = -p α = -pa	τ _{a⁻} = -p α = -p(a-d)	τ _{a⁻} = -p α = -p(a+d)
τ _a	α = α ⁺ - α ⁻ = [a + d (q-p)] = [a + d(1-2p)]	τ _a = α ⁺ - α ⁻ = [a + d (s-r)] = [a + d(1-2r)]	τ _a = α ⁺ - α ⁻ = a	τ _a = α ⁺ - α ⁻ = a-d	τ _a = α ⁺ - α ⁻ = a+d

Tabla 4.7.

Efectos medios de un gen y efecto promedio de sustitución de un gen de una población en equilibrio de Hardy-Weinberg utilizando diferentes probadores.

4.7. Habilidad combinatoria

Sprague y Tatum (1942) acuñaron el término habilidad (aptitud) combinatoria (*combining ability*) para describir al comportamiento de los cruzamientos entre líneas endocriadas de maíz. El término habilidad o aptitud combinatoria general (ACG_i) hace referencia al desvío entre la media de los cruzamientos

de la línea i -ésima con un número variable de otras líneas, y la media general a través de todos los cruzamientos. La habilidad o aptitud combinatoria específica (ACE_{ij}) se calcula como el desvío entre la media de un cruzamiento entre las líneas i y j y la suma de la media general y las habilidades combinatorias generales de ambas líneas progenitoras del cruzamiento. Estas definiciones pueden expresarse mediante el modelo que sigue:

$$M_{F1(ij)} = \mu + ACG_{(i)} + ACG_{(j)} + ACE_{(ij)} \quad [4.15],$$

donde

μ es la media general de los cruzamientos entre un grupo de n líneas,

$M_{F1(ij)}$ es la media observada para el cruzamiento o híbrido F_1 entre las líneas i y j ,

$ACG_{(i)}$ es la habilidad o aptitud combinatoria general de la línea progenitora i ,

$ACG_{(j)}$ es la habilidad o aptitud combinatoria general de la línea progenitora j ,

$ACE_{(ij)}$ es la habilidad o aptitud combinatoria específica del cruzamiento i x j .

En un sentido amplio y sin tener en cuenta la descomposición en efectos de habilidad combinatoria general y específica, Hallauer y Miranda Filho (1988) describen la habilidad combinatoria de un grupo de genotipos pertenecientes a una población en equilibrio de Hardy-Weinberg, en cruzamiento con otra población probadora (Tabla 4.8).

Genotipo	Frec.	Frec. alelo favorable	Media de cruzamiento de prueba	Habilidad combinatoria	Habilidad combinatoria según frecuencias génicas del probador		
					$r=0$	$r=1/2$	$r=1$
L_2L_2	q^2	0	T_{22}	$T_{22} - T_{..}$	$-p(a+d)$	$-pa$	$-p(a-d)$
L_1L_2	$2pq$	$1/2$	T_{12}	$T_{12} - T_{..}$	$(1/2-p)(a+d)$	$(1/2-p)(a)$	$(1/2-p)(a-d)$
L_1L_1	p^2	1	T_{11}	$T_{11} - T_{..}$	$(1-p)(a+d)$	$(1-p)(a)$	$(1-p)(a-d)$
Media		P	T..	T..			

Tabla 4.8.

Habilidad combinatoria general de individuos extraídos de una población en equilibrio de Hardy-Weinberg respecto de probadores con diferentes frecuencias génicas.

Previamente definimos como efecto de sustitución de un gen en cruzamientos de prueba (utilizando como probadora a otra población) a la expresión:

$$\alpha_T = a + d(s - r) = a + d[(1 - r) - r] = [a + d(1 - 2r)]$$

En consecuencia la habilidad combinatoria general es el producto del efecto de sustitución de un gen en cruzamientos de prueba multiplicado por las diferencias entre las frecuencias génicas para el alelo favorable ($p_{ij} - p$) en el genotipo evaluado (p_{ij}) y para el promedio de la población evaluada (p) (Tabla 4.8).

La habilidad combinatoria (C) para cada genotipo de la población será:

$$C_{ij} = T_{ij} - T_{..} = (p_{ij} - p) [a + (1 - 2r)d]$$

$$C_{22} = T_{22} - T_{..} = (0 - p)[a + (1 - 2r)d] = -p[a + (1 - 2r)d]$$

$$C_{12} = T_{12} - T_{..} = (0,5 - p)[a + (1 - 2r)d]$$

$$C_{11} = T_{11} - T_{..} = (1 - p)[a + (1 - 2r)d]$$

Obsérvese que las casillas destacadas en negrita en las Tablas 4.7 y 4.8 se corresponden entre sí, solo que se presentan transpuestas.

Del análisis de la Tabla 4.8 y los comentarios anteriores, se concluye que en el caso de que $r = 1$ en el probador (el alelo favorable está fijado) y la dominancia sea completa ($a = d$), la habilidad combinatoria será igual a cero en todos los cruzamientos de prueba y el probador no puede, consecuentemente, discriminar entre los genotipos. Volveremos sobre esta cuestión en el capítulo de variancias.

La F_1 del cruzamiento interpoblacional no se encuentra en equilibrio de Hardy-Weinberg y las medias se modificarán ya sea por autofecundación como por apareamiento aleatorio. En cualquier caso, en la medida que exista dominancia unidireccional o bidireccional neta que favorezca la expresión del carácter, la media poblacional resultante será menor que M_{F_1} por efecto de la depresión por endocría. Luego de una generación de apareamiento aleatorio se alcanzará el equilibrio de Hardy-Weinberg, donde las frecuencias genotípicas pueden deducirse fácilmente a partir de las nuevas frecuencias génicas $p' = (1/2)(p + r)$ y $q' = (1/2)(q + s)$. En el caso de la formación de compuestos o poblaciones complejas a partir de n poblaciones parentales, las frecuencias génicas para los alelos favorables y desfavorables serán (si las contribuciones de las poblaciones son equivalentes):

$$p' = (1/n)(p_1 + p_2 + p_3 + \dots + p_n) \text{ y } q' = 1 - p', \quad [4.16]$$

donde p_i son las frecuencias del alelo favorable en cada una de las poblaciones parentales.

4.7.1. Modelo de cruzamientos de prueba de Melchinger

Melchinger (1988) presentó dos modelos que describen el comportamiento medio de una generación de cruzamientos de prueba de una población de cría resultante del híbrido entre dos líneas pertenecientes a un mismo grupo heterótico ($I_1 \times I_2$), empleando una línea probadora perteneciente a un grupo heterótico opuesto ($^T I$). Se trata de un análisis de medias de los cruzamientos de prueba versus $^T I$ de las generaciones F_1 , F_2 y sendas retrocruzas a una y otra de las líneas progenitoras de la población de cría, es decir, RCI_1 y RCI_2 . Adicionalmente se avanzan generaciones desde la F_2 hasta F_n mediante autofecundaciones y por apareamiento aleatorio. Ambos modelos difieren entre sí en los supuestos sobre ligamiento y epistasis. El modelo I asume presencia de ligamiento y ausencia de epistasis y el modelo II considera ausencia de ligamiento y presencia de epistasis.

En el modelo I de Melchinger:

$$Y = {}^T M + x({}^T d),$$

donde

Y es la media del cruzamiento de prueba de una generación,

${}^T M$ es la media del cruzamiento de prueba de $^T I$ por la población F_2 derivada del cruzamiento entre $I_1 \times I_2$, en equilibrio gamético,

$$({}^T d) = \sum \theta_i ({}^T d_i),$$

siendo $\theta_i = +1$ si la línea I_1 lleva en el locus i -ésimo el alelo favorable y -1 si lleva el alelo desfavorable.

$$({}^T d_i) = \frac{1}{2} \alpha_i,$$

donde α_i es el efecto promedio de sustitución en el locus i -ésimo en el cruzamiento de prueba de la población F_2 .

Finalmente x es un coeficiente que depende de la generación que se analiza.

El modelo II de Melchinger es:

$$Y = {}^T M + x({}^T d) + x^2({}^T ie),$$

donde el segundo término que se incorpora es

$$({}^T ie) = \sum \theta_i \theta_j ({}^T ie_{ij}), \text{ para } i < j \text{ y finalmente}$$

$({}^T ie_{jk}) =$ es el efecto de interacción epistática aditiva x aditiva del i -ésimo locus con el j -ésimo locus.

Analizando cruzamientos de prueba en maíz mediante los modelos planteados por Melchinger, Lamkey *et al.* (1995) demostraron que tanto el modelo I como el II mostraron faltas de ajustes significativas. Sin embargo con el modelo II se detectaron efectos epistáticos significativos entre loci no ligados que explicaban un tercio del ajuste del modelo. La falta de ajuste de ambos modelos sugeriría la existencia de otros tipos de interacciones (dominante x dominante, aditiva x dominante, etc.) que contribuyen a explicar la variancia en el comportamiento de los cruzamientos de prueba. Como señalan Lamkey *et al.* (1995), estos resultados que indican la importancia de los efectos epistáticos plantearía limitaciones al éxito de esquemas de selección asistida por marcadores moleculares cuyo objetivo es la acumulación de alelos favorables en un progenitor.

4.8. Estimación y predicción de la media esperada de híbridos, variedades sintéticas y compuestos

Una de las ventajas de organizar el germoplasma de un programa de mejoramiento en grupos genéticamente distintivos es que el número de híbridos para evaluar será considerablemente menor que en el caso de programas con germoplasma no estructurado en grupos. La formación de grupos heteróticos permite concentrar los esfuerzos en la predicción, formación, evaluación y selección de híbridos con mayor probabilidad de maximizar la heterosis.

Cuando existen dos grupos heteróticos que muestran buen comportamiento general en sus cruzamientos y se dispone de n_1 líneas parentales de un grupo y n_2 líneas parentales del segundo, el número de híbridos simples posibles será:

$$N_{HS} = n_1 n_2, \quad [4.17]$$

El número de híbridos triples (cruzamiento de todos los híbridos simples posibles por la línea restante) considerando dos grupos heteróticos será:

$$N_{HT} = n_1 C_{n_2}^2 \text{ o bien } N_{HT} = n_2 C_{n_1}^2, \quad [4.18]$$

si el grupo de líneas elegido para hacer el híbrido simple pertenece al grupo 1 o al grupo 2, respectivamente.

El número de híbridos dobles (cruzamientos entre los híbridos simples posibles) será el producto del número combinatorio de n_1 líneas tomadas de a 2 multiplicado por el número combinatorio de n_2 líneas tomadas de a 2. De tal manera, el número total de híbridos dobles será:

$$N_{HD} = C_{n_1}^2 C_{n_2}^2 = [n_1! / (n_1 - 2)! 2!] \times [n_2! / (n_2 - 2)! 2!] \text{ híbridos dobles} \quad [4.19]$$


4.8.1. Predictores BLUP (*Best Linear Unbiased Predictors*) del valor genotípico

En esta sección describiremos la aplicación de los modelos mixtos (Anexo) para obtener predictores del comportamiento de genotipos. Una situación donde ello resulta ventajoso en programas de mejoramiento es cuando los genotipos se evalúan en experimentos replicados en una serie de ambientes y los datos resultan desbalanceados, es decir, que la cantidad de información disponible de cada genotipo no es uniforme. Otra situación es en programas de desarrollo de híbridos o variedades donde el número potencial de híbridos o variedades excede las posibilidades factibles de evaluación a campo. La aplicación de modelos mixtos permite obtener predictores del comportamiento de híbridos simples y extenderlo a la predicción de híbridos de tres líneas y dobles (Bernardo, 2010). Adicionalmente, pueden obtenerse predictores de la habilidad combinatoria general, habilidad combinatoria específica y de la interacción de ambas con el ambiente (Boca y Cantet, 2004; Eyherabide et al., 2016).

4.8.1.1. Predicción BLUP y GBLUP del comportamiento de híbridos simples

A partir de la matriz de parentesco entre una serie de líneas progenitoras de híbridos y de evaluaciones fenotípicas de un número determinado de ellos, se busca obtener predictores de comportamiento de híbridos simples para los que no se dispone de información fenotípica. Como se mencionó anteriormente, frecuentemente los programas de desarrollo de híbridos tienen estructurado o clasificado su germoplasma en grupos heteróticos, de modo que al cruzar genotipos de uno y otro entre sí, hay mayores probabilidades de obtener híbridos de buen comportamiento. Suponiendo que se partiera de dos grupos de líneas, clasificadas según los grupos heteróticos A y B, se requiere conocer la matriz de coeficientes de coancestría entre las líneas del grupo A y entre las líneas del grupo B y disponer de información de un grupo parcial de los híbridos posibles A x B. Mediante el empleo de las ecuaciones de los modelos mixtos (Anexo) se estiman la variancia del error y la genotípica, los efectos fijos correspondientes a los parámetros ambientales y los predictores de la habilidad combinatoria general de cada línea y de la habilidad combinatoria específica de cada cruzamiento.

En caso de disponerse de datos de marcadores moleculares en alta densidad de las líneas progenitoras pueden obtenerse aún mejores predictores genómicos del comportamiento de híbridos (GBLUPs). La calidad de la predicción mejora en la medida que se combinen en el modelo mixto la información de ambas matrices, esto es la de relaciones entre líneas progenitoras basadas en pedigrís y la de relaciones genómicas a partir de los datos de marcadores moleculares.



En programas de desarrollo de híbridos que generan un número significativo de líneas parentales es prácticamente imposible hacer todos los cruzamientos posibles y efectuar la evaluación de todos los híbridos resultantes. Contar con un modelo de predicción del comportamiento de híbridos permite desestimar la realización y evaluación a campo de una fracción importante del total de híbridos posibles y concentrar los recursos para evaluar los restantes.

Tratándose de híbridos, su predicción del comportamiento tiene en cuenta las predicciones de los efectos de habilidad combinatoria general de las líneas parentales y las predicciones de los efectos de habilidad combinatoria específica de los cruzamientos. La predicción del comportamiento de los cruzamientos puede realizarse utilizando los modelos que se emplean para estimar el valor de mejora genómico y efectuar selección genómica (Vélez-Torres *et al.*, 2018). Technow *et al.* (2014) utilizaron un modelo GBLUP para predecir el comportamiento de híbridos en un rango ambiental tomando en cuenta la matriz de relaciones genómicas entre las líneas parentales utilizando marcadores moleculares en alta densidad. La calidad de las predicciones del comportamiento de híbridos depende de la heredabilidad del carácter evaluado, la importancia de las interacciones genotipo x ambiente para el carácter, la densidad de marcadores, las características de las poblaciones de entrenamiento y de predicción, y el modelo de predicción utilizado (Vélez-Torres *et al.*, 2018). La población de entrenamiento provee los datos de marcadores moleculares de las líneas progenitoras y los datos fenotípicos de sus cruzamientos.

El modelo predictivo lineal presentado por Technow *et al.* (2014) consiste en:

$$y = X\beta + Z_A g_A + Z_B g_B + Z_{AxB} h_{AxB} + e$$

donde:

y es vector de predicción del comportamiento del híbrido resultante del cruzamiento de la f -ésima línea hembra por la m -ésima línea macho en cada ambiente de evaluación,

X es la matriz de diseño correspondiente a los ambientes,

β es el vector de los efectos de cada ambiente, considerados fijos,

Z_A es la matriz de diseño correspondiente a los efectos de habilidad combinatoria general de las líneas hembras,

Z_B es la matriz de diseño correspondiente a los efectos de habilidad combinatoria general de las líneas machos,

$Z_{A \times B}$ es la matriz de diseño correspondiente a los efectos de habilidad combinatoria específica de los híbridos $f \times m$,

g_A es el vector de efectos aleatorios de habilidad combinatoria general de las líneas hembras o del grupo heterótico A,

g_B es el vector de efectos aleatorios de habilidad combinatoria general de las líneas machos o del grupo heterótico B,

$h_{A \times B}$ es el vector de efectos aleatorios de habilidad combinatoria específica de los cruzamientos A x B, y

e es el vector de residuales o efectos no considerados por el modelo.

Los vectores de efectos de habilidad combinatoria ($g_A, g_B, h_{A \times B}$) se distribuyen todos con media 0 y sus respectivas variancias son $G_A \times \sigma_{HCGA}^2, G_B \times \sigma_{HCGB}^2$ y $H_{A \times B} \times \sigma_{HCE}^2$, donde:

$\sigma_{HCGA}^2, \sigma_{HCGB}^2$ y σ_{HCE}^2 son las variancias de los efectos de habilidad combinatoria general de las líneas del grupo A, de los efectos de habilidad combinatoria de las líneas B, y la variancia de los efectos de habilidad combinatoria específica y

G_A y G_B son las matrices genómicas de parentesco entre líneas de A y entre líneas de B obtenidas a partir de los marcadores moleculares siguiendo a VanRaden (2008) como:

$$G_A = W_A W_A' / n_A \quad y$$

$$G_B = W_B W_B' / n_B,$$

donde W_A y W_B son sendas matrices de marcadores centradas y estandarizadas, y luego divididas por el número de marcadores empleados, tanto para las líneas hembras y líneas machos. Los elementos de estas matrices correspondientes al progenitor k-ésimo y el marcador n-ésimo se computan como:

$$w_{kn} = [(x_{kn} - 2p_n) / 2(p_n(1-p_n))]^{1/2}$$

donde:

p_n es la frecuencia alélica del marcador n-ésimo y x_{kn} es la codificación que se utilice para el número de alelos de referencia.

La matriz H_{AxB} es la matriz producto Kronecker (ver Anexo) entre las matrices genómicas de parentesco:

$$H_{AxB} = G_A \otimes G_B$$

El vector de los residuales, e , se distribuye con media cero y su variancia será $R\sigma_R^2$, donde σ_R^2 es la variancia de los residuales, y R es una matriz diagonal cuyos elementos de la diagonal son $(b)^{-1}$, donde b es el número de repeticiones en el ambiente correspondiente.

Acosta-Pech *et al.* (2017) encontraron que la habilidad predictiva del modelo podía mejorarse incorporando la consideración de la interacción genotipo x ambiente.

4.8.1.2. Media esperada de híbridos de tres y cuatro líneas

Los primeros métodos de predicción del comportamiento de híbridos dobles se generaron a partir de supuestos sobre la diferente importancia relativa de cada tipo de acción génica y estaban basados en el comportamiento de los híbridos simples obtenibles de la combinación de sus líneas parentales (Jenkins, 1934). Otros modelos posteriores de predicción (Cockerham, 1967) también se basaron en el comportamiento de híbridos simples y adicionalmente de híbridos triples (Eberhart, 1964). Cockerham (1967) y Otsuka *et al.* (1972) propusieron un predictor óptimo del comportamiento de los híbridos basado en los efectos de habilidad combinatoria general y específica estimados a partir de cruzamientos dialélicos entre las líneas progenitoras. A manera de ejemplo, la predicción del comportamiento del híbrido de tres líneas, $(A \times B) \times C$, será la siguiente:

$$MP_{ABC} = m + \frac{1}{2} (ACG_A + ACG_B) + ACG_C + \lambda \frac{1}{2} (ACE_{AC} + ACE_{BC}), \quad [4.20]$$

donde λ es un factor de ponderación (variable entre 0 y 1) de los efectos de habilidad combinatoria específica de las combinaciones no parentales (ACE_{AC} , ACE_{BC}) del híbrido de tres líneas que se desea predecir. ACG_A , ACG_B y ACG_C son las habilidades combinatorias generales de las líneas A , B y C , respectivamente.

En general, los mejores modelos de predicción tuvieron eficiencias similares al método B de Jenkins. Este predice el comportamiento de un híbrido doble como el promedio de los cuatro híbridos simples posibles entre las cuatro líneas parentales, combinadas en híbridos simples no parentales. Cuando las líneas parentales son altamente seleccionadas, el método A de Jenkins, basado en el promedio de todos los híbridos simples posibles de formar a partir de cuatro líneas progenitoras, también muestra alta eficien-

cia predictiva. Así, el comportamiento del híbrido doble (AB) x (CD) puede predecirse (según Jenkins B y Jenkins A, respectivamente) como:

$$PM_{AB,CD} = \frac{1}{4} [M_{AC} + M_{AD} + M_{BC} + M_{BD}], \quad [4.21]$$

$$PM_{AB,CD} = \frac{1}{6} [M_{AC} + M_{AD} + M_{BC} + M_{BD} + M_{AB} + M_{CD}], \quad [4.22]$$

Siguiendo similares razonamientos, el procedimiento Jenkins B se extendió a la predicción del comportamiento de híbridos de tres líneas. En el caso del híbrido de tres líneas (AB) x C la predicción será:

$$PM_{AB,C} = \frac{1}{4} (M_{AC} + M_{BC}), \quad [4.23]$$

Frec. Gaméticas en Híbrido simple progenitor $L_1 \times L_2$	Frec. Gaméticas en Línea Progenitora L_3	
	$p=1$	$q=0$
Frecuencias genotípicas x Valores Genotípicos		
$p=0.5$	$p(a) = \frac{1}{2} a$	0
$q=0.5$	$q(d_{12}) = \frac{1}{2} d_{12}$	0
Media F_1	$\frac{1}{2} (a + d_{12})$	

Tabla 4.9.

Media de híbridos de tres líneas en función de la constitución genotípica de sus líneas progenitoras considerando un locus con dos alelos.

La media esperada del híbrido de tres líneas (AB) x C, asumiendo que la línea progenitora macho posee un alelo diferente al que lleva el híbrido parental hembra AB será como se indica en la Tabla 4.10. Nótese que se capitalizan los desvíos de dominancia entre las combinaciones de posibles de alelos provenientes de las tres líneas.

Frec. Gaméticas en Híbrido simple progenitor $L_1 \times L_2$	Frec. Gaméticas en Línea Progenitora L_3	
	$r=1$	$s=0$
Frecuencias genotípicas x Valores Genotípicos		
$p=0.5$	$p(d_{13}) = \frac{1}{2} d_{13}$	0
$q=0.5$	$q(d_{23}) = \frac{1}{2} d_{23}$	0
Media F_1	$\frac{1}{2} (d_{13} + d_{23})$	

Tabla 4.10.

Media de híbridos de tres líneas en función de la constitución genotípica de las líneas involucradas considerando un locus con alelos múltiples.

4.8.2. Media esperada de variedades sintéticas y de compuestos

4.8.2.1. Variedades sintéticas

Las variedades sintéticas son poblaciones que se obtienen por el inter cruzamiento de líneas parcial o totalmente homocigotas. Posteriormente a la realización de los cruzamientos, la población se conduce hacia el equilibrio gamético mediante apareamiento aleatorio y en condiciones de aislamiento para prevenir eventuales contaminaciones con polen extraño. Dentro de esta categoría de variedades se encuentran las poblaciones resultantes de la recombinación de familias endocriadas seleccionadas en cada ciclo de selección recurrente (ver más adelante). Cuando se dispone de n líneas endocriadas puede generarse un número potencial de variedades sintéticas igual a:

$$N_{VS} = 2^n - (n + 1), \quad [4.24]$$

La media de una variedad sintética resultante de la recombinación o inter cruzamiento entre n líneas puede predecirse a partir de la fórmula de Wright (Wright, 1922) y Kinman y Sprague (1945), citados por Hallauer y Miranda Filho (1988):

$$MP_{VS} = M_{HSn} - (M_{HSn} - M_n)/n, \quad [4.25]$$

donde M_{HSn} es el comportamiento promedio de todos los híbridos simples posibles entre las n líneas y M_n es el promedio de comportamiento de las líneas parentales. Cuanto menor sea la superioridad de los híbridos simples respecto de las líneas parentales mayor tenderá a ser la predicción del comportamiento promedio de la variedad sintética. Cuanto mayor sea el número de líneas parentales, la media esperada de la variedad sintética tenderá a ser mayor. En otras palabras, a bajos niveles de heterosis y mayor número de líneas progenitoras de las sintéticas, mayor tenderá a ser su media esperada.

Busbice (1970) propuso una fórmula general de estimación que tiene en cuenta el coeficiente de endocría de las diferentes generaciones de la variedad sintética, que aplicado al ejemplo anterior será

$$MP_{VS2} = M_{HSn} + [(F_0 - F_2)/(F_0 - F_1)](M_{HSn} - M_n), \quad [4.26]$$

en la que la magnitud del parámetro F_i es función del número de progenitores de la variedad sintética y de la tasa de autofecundación natural.

4.8.2.2. Compuestos

Los compuestos son poblaciones de amplia base genética que se obtienen por inter cruzamiento de dos o más poblaciones heterogéneas. Posteriormente al inter cruzamiento, la multiplicación del compuesto se hace mediante apareamiento aleatorio y en condiciones de aislamiento para prevenir eventuales contaminaciones con polen extraño.

A partir de n poblaciones pueden generarse un número potencial de compuestos (N_c) según una expresión esencialmente similar a [4.24]:

$$N_c = 2^n - (n + 1), \quad [4.27]$$

Eberhart *et al.* (1967) presentaron una fórmula de predicción de compuestos en condiciones de apareamiento aleatorio similar a la fórmula de Wright:

$$MP_c = M_{P \times P} - (M_{P \times P} - M_p)/n, \quad [4.28]$$

donde $M_{P \times P}$ es el comportamiento promedio de los cruzamientos interpoblacionales posibles, M_p es el comportamiento promedio de las poblaciones parentales, y n el número de poblaciones que formarán el compuesto.

Vencovsky (1980) describe la media esperada de un compuesto obtenido a partir de la recombinación de n poblaciones cuando además del comportamiento medio de las poblaciones parentales se conoce la heterosis promedio (H_M) entre ellas:

$$MP_c = M_{P \times P} + [(n - 1)/n]H_M, \quad [4.29]$$

La heterosis promedio es la media de la heterosis que presentan los $[n(n-1)/2]$ cruzamientos posibles entre las n poblaciones parentales. Cuanto mayor sea el número de poblaciones incorporadas al compuesto, mayor será la proporción esperada de la heterosis promedio que contribuye al comportamiento promedio del compuesto.

4.8.3. Modelo general de Vencovsky y Barriga (1992) para predecir medias de híbridos y de compuestos derivados de cruzamientos interpoblacionales

El cruzamiento F_1 (Z) entre dos parentales genéticamente diferentes, (X) e (Y), puede expresarse simbólicamente mediante la expresión general:

$$(Z) = (X).(Y), \quad [4.30]$$

Para predecir la media de cada cruzamiento se reemplazan los términos parentales, (X) e (Y) por sus equivalentes. Si por ejemplo (Z) fuera un híbri-

do doble (AB) x (CD), producto del cruzamiento entre los híbridos simples (AB) y (CD), X e Y corresponden a las contribuciones de cada uno de sus padres, en una proporción equivalente al 50 %. De tal modo:

$$(X) = (\frac{1}{2})A + (\frac{1}{2})B,$$

$$(Y) = (\frac{1}{2})C + (\frac{1}{2})D$$

Por lo tanto:

$$(Z) = [(\frac{1}{2})A + (\frac{1}{2})B] \cdot [(\frac{1}{2})C + (\frac{1}{2})D] = \frac{1}{4} [AC + AD + BC + BD]$$

Obsérvese que según esta expresión, la media estimada de un híbrido doble es análoga al promedio de los cuatro híbridos simples posibles no parentales, tal lo que indica el método de predicción B de Jenkins.

Expresando (Z) conforme al modelo inicial basado en la media de los dos homocigotas para el alelo favorable y el desfavorable, y el desvío de dominancia (ver Capítulo 3), sabemos que las medias de los híbridos AC, AD, BC y BD serán, respectivamente:

$$AC = m + d_{AC}$$

$$AD = m + d_{AD}$$

$$BC = m + d_{BC}$$

$$BD = m + d_{BD}$$

$$Z = m + \frac{1}{4}[d_{AC} + d_{AD} + d_{BC} + d_{BD}]$$

En el caso de querer estimar el comportamiento de la F_2 del híbrido doble (AB) x (CD), ello es equivalente a:

$$Z = [(X).(Y)] \cdot [(X).(Y)] = [(AB) \times (CD)] \cdot [(AB) \times (CD)] = [(AB) \times (CD)]^2 =$$

$Z = [(\frac{1}{4})A + (\frac{1}{4})B + (\frac{1}{4})C + (\frac{1}{4})D]^2$, por cuanto la contribución genética de las líneas parentales al híbrido doble es 1/4 para cada una de las líneas parentales A, B, C, D.

$$Z = (1/16)(A + B + C + D + 2AB + 2AC + 2AD + 2BC + 2BD + 2CD)$$

Reemplazando estas expresiones por sus componentes genéticos:

$$\begin{aligned} Z = (1/16) [m + a_A + m + a_B + m + a_C + m + a_D + 2(m + d_{AB}) + \\ + 2(m + d_{AC}) + 2(m + d_{AD}) + 2(m + d_{BC}) + 2(m + d_{BD}) + \\ + 2(m + d_{CD})] = \end{aligned}$$

$$Z = m + (1/16)[a_A + a_B + a_C + a_D] + (1/8) [d_{AB} + d_{AC} + d_{AD} + d_{BC} + d_{BD} + d_{CD}]$$

Observando esta última expresión se advierte que en la F_2 de un híbrido doble se pierde la mitad de los efectos de dominancia que manifiesta el híbrido doble F_1 y esa contribución es compensada parcialmente por los efectos aditivos de las líneas progenitoras. Debe recordarse que en el caso de la F_2 de un híbrido simple, se pierde $1/2$ de los efectos de dominancia que se presentan en el híbrido F_1 .

Si Z fuera un híbrido entre dos poblaciones con frecuencias génicas (p, q) y (r, s), respectivamente:

$$\begin{aligned} Z &= (X).(Y) \\ Z &= \{1/2[a(p - q) + 2pqd]\} . \{1/2 [a(r - s) + 2rsd]\} = \\ Z &= [(1/2)a(p - q) + pqd] [(1/2)a(r - s) + rsd] = \\ &= (1/4)a^2(p-q)(r-s) + (1/2)ad(p-q)rs + (1/2)ad(r - s)pq + d^2rspq... \\ &= a(r + p - 1) + d(p - r) \end{aligned}$$

El modelo general de Vencovsky puede aplicarse también para predecir el comportamiento del compuesto obtenible por el cruzamiento entre dos o más poblaciones luego de las recombinaciones necesarias hasta obtener el equilibrio de ligamiento:

$Z = (X).(Y) = (X).(X)$, por cuanto nos estamos refiriendo al comportamiento de una única población en equilibrio de H_w resultante del cruzamiento entre dos o más (n) poblaciones.

$$\begin{aligned} Z &= (X)^2 = [(1/n)(P_1 + P_2 + \dots + P_n)]^2 = (1/n^2)[P_1 + P_2 + \dots + P_n + \\ &+ 2P_{12} + 2P_{13} + \dots + 2P_{1n} + 2P_{23} + \dots + 2P_{2n} + \dots + 2P_{(n-1)n}] = \\ Z &= (1/n) P. + [1 - (1/n) C.], \end{aligned}$$

siendo P. el comportamiento promedio de las poblaciones progenitoras y C. el comportamiento promedio de los cruzamientos interpoblacionales entre las n poblaciones.

Si se trata de $n=2$,

$$C = (Z) = (1/2)[P_1 + P_2] + [1 - (1/2)P_{12}]$$

Cuando se quiere formar un compuesto a partir del cruzamiento de varias poblaciones, pero se desea que la contribución de las poblaciones sea desigual, podemos aplicar el modelo general de Vencovsky. Supongamos que

disponemos de cuatro poblaciones, P_1 , P_2 , P_3 y P_4 y que deseamos que la proporción genética de P_4 en el compuesto sea del 50 %, el comportamiento promedio del compuesto puede predecirse a partir de:

$$PM_c = (Z) = (X)^2 = \left\{ \left(\frac{1}{2} \right) \left[\left(\frac{1}{3} \right) P_1 + \left(\frac{1}{3} \right) P_2 + \left(\frac{1}{3} \right) P_3 \right] + \left(\frac{1}{2} \right) P_4 \right\}^2$$

Los componentes de la predicción del comportamiento del compuesto pueden calcularse a partir del cuerpo de la Tabla 4.11, donde los términos P_i y P_{ij} corresponden a las medias de comportamiento de las poblaciones parentales y las de los híbridos interpopulacionales, respectivamente.

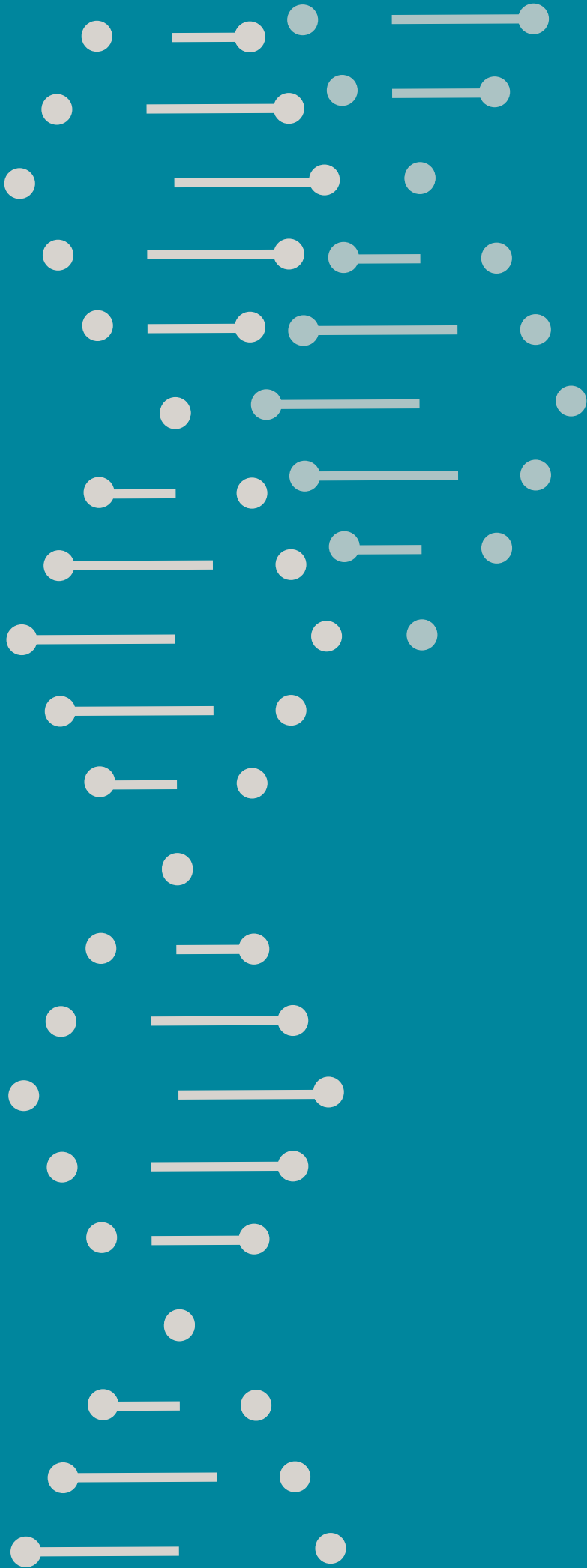
Contribuciones de las poblaciones parentales	$(1/6)P_1$	$(1/6)P_2$	$(1/6)P_3$	$(1/2)P_4$
$(1/6)P_1$	$(1/36)P_1$	$(1/36)P_{21}$	$(1/36)P_{31}$	$(1/12)P_{14}$
$(1/6)P_2$	$(1/36)P_{12}$	$(1/36)P_2$	$(1/36)P_{32}$	$(1/12)P_{24}$
$(1/6)P_3$	$(1/36)P_{13}$	$(1/36)P_{23}$	$(1/36)P_3$	$(1/12)P_{34}$
$(1/2)P_4$	$(1/12)P_{14}$	$(1/12)P_{24}$	$(1/12)P_{34}$	$(1/4)P_4$

Tabla 4.11.

Coeficientes de contribución de las medias parentales al comportamiento de un compuesto en función de la participación de cada población parental. Se asume que las poblaciones 1, 2 y 3 hacen contribuciones equivalentes al 50 % de la composición genética del compuesto y la población 4 contribuye con el 50 % restante.

En resumen, la media de este compuesto será:

$$PM_c = \left(\frac{1}{4} \right) P_4 + \left(\frac{1}{36} \right) (P_1 + P_2 + P_3) + \frac{1}{6} (P_{14} + P_{24} + P_{34}) + \left(\frac{1}{18} \right) (P_{12} + P_{13} + P_{23})$$





Variabilidad poblacional

5.1. Variancia fenotípica y sus componentes

En el Capítulo 3 se definieron los valores y efectos que resultan en la expresión o valores fenotípicos para un carácter cuantitativo. Estos últimos son particulares de cada individuo evaluado en determinado ambiente. Las variaciones entre individuos estudiadas a nivel poblacional pueden cuantificarse mediante estadísticas de dispersión, la más común es la variancia fenotípica. Esta, a su vez, resulta del aporte de las variancias correspondientes de cada uno de los valores y efectos que causan el valor fenotípico. En otras palabras, la variancia fenotípica mide la dispersión entre los valores fenotípicos de los individuos que integran una población genética y que resulta de la contribución de la variancia de los valores genotípicos (variancia genotípica) y de la variancia de los desvíos ambientales (variancia ambiental).

Si nos remitimos a la expresión:

$$P = G + E \quad [3.1]$$

la variabilidad fenotípica entre los individuos puede describirse como:

$$V_P = V_{(G+E)} \quad [5.1]$$

Como el miembro de la derecha de la igualdad es una función lineal, la ecuación [5.1] resulta equivalente a:

$$V_P = V_G + V_E + 2Cov_{GE} \quad [5.2]$$

Toda vez que exista independencia (o ausencia de correlación estadística) entre los valores genotípicos y los desvíos ambientales, es decir, que cualquier genotipo en evaluación tiene la misma probabilidad que otro de ser asignado a cualquier condición ambiental, el término de covarianza se iguala a cero. La aleatorización en los experimentos permite aproximarnos a esta situación de independencia, de tal modo que la expresión [5.2] puede simplificarse como:

$$V_P = V_G + V_E \quad [5.3]$$

donde:

V_P es la variancia fenotípica,


V_G es la variancia genotípica y

V_E es la variancia ambiental.

5.1.1. Abordajes en el estudio de la variabilidad fenotípica

Al estudiar la variabilidad de una población genética nuestro interés no se focaliza simplemente en cuantificar la variancia fenotípica, sino que el mayor interés está en determinar la magnitud e importancia relativa de sus componentes de origen genotípico y ambiental, que son componentes causales de la variabilidad fenotípica. A su vez, el componente de variabilidad genotípica puede partitionarse en subcomponentes asociados a los diferentes modos de acción génica: aditividad, dominancia y epistasis.

Cuando una población genética en estudio es homogénea (todos los individuos que la integran son genotípicamente idénticos), como por ejemplo en el caso de una muestra de individuos de una línea completamente homocigota o de individuos híbridos F_1 , producto del cruzamiento entre dos líneas progenitoras completamente homocigotas, la variancia fenotípica



que podemos medir en la población provee una estimación de la variancia ambiental o de la variabilidad de los desvíos ambientales, ya que no existe variabilidad de origen genético en la muestra.

En poblaciones que son heterogéneas debido a que existen diferencias genotípicas entre los individuos que la componen, estudiar su variabilidad fenotípica resulta más compleja por cuanto será necesario un procedimiento experimental adecuado para estimar ambos componentes, genético y ambiental.

A los fines de su análisis experimental, una población genética puede estructurarse en grupos de individuos emparentados (familias) y la variancia fenotípica dividirse, entonces, en dos componentes: la variancia fenotípica entre familias y la variancia fenotípica dentro de familias. A estos componentes se los denomina componentes observacionales de la variancia fenotípica. Ello puede expresarse como:

$$V_P = V_{PEF} + V_{PDF}$$

donde:

V_{PEF} es la variancia fenotípica entre familias y

V_{PDF} es la variancia fenotípica dentro de familias.

Tanto la variancia fenotípica entre familias como la variancia fenotípica dentro de familias responden, a su vez, a sus respectivos componentes causales. De tal modo que en las variancias fenotípicas entre familias y dentro de familias existirán sendos subcomponentes debidos a factores genéticos y ambientales.

Los diseños experimentales que se emplean para estimar los componentes causales de la variancia fenotípica se denominan genéricamente diseños genéticos y su capacidad para estimarlos y el nivel de precisión de los estimadores es proporcional a la complejidad del diseño genético. No obstante ello, a los fines prácticos y para caracteres de herencia medianamente conocida, puede resultar suficiente la aplicación de diseños sencillos como los que se utilizan rutinariamente en los programas de mejoramiento para evaluar genotipos. En otros casos hay que emplear diseños genéticos que requieren la evaluación de combinaciones de tipos de familias generadas según los objetivos que se persigan.

La variabilidad fenotípica, entonces, puede partitionarse simultáneamente en sus componentes causales y observacionales. En estos esquemas debemos recordar que la suma de los componentes observacionales (entre y dentro de familias) de la variabilidad genotípica proveerá una estimación de la va-

riabilidad genotípica total de la población. La variabilidad ambiental también afecta a los dos componentes observacionales (entre y dentro de familias).

En consecuencia, lo indicado más arriba puede expresarse mediante la siguiente ecuación:

$$V_P = V_G + V_E = (V_{GEF} + V_{GDF}) + (V_E + V_{EW}) \quad [5.4]$$

donde V_{GEF} y V_{GDF} son los componentes de variancia genotípica entre familias y dentro de familias, respectivamente. V_E y V_{EW} corresponden a la variancia ambiental entre familias y dentro de familias.

Dependiendo del tipo de familias que estemos utilizando en el diseño genético, los componentes causales de la variancia genotípica que se expresan entre y dentro de familias serán diferentes. En lo sucesivo, cuando hagamos mención a la variancia de un tipo de familia nos estaremos refiriendo a la variancia entre familias.

En términos estadísticos y de diseño experimental, las familias son aleatoriamente asignadas a unidades experimentales en cada repetición y la variancia fenotípica de familias responde a $V_{GEF} + V_e$, donde V_e es la variancia ambiental entre parcelas y V_{GEF} es la variabilidad genotípica que se expresa entre unidades experimentales asignadas a distintas familias.

$$V_{PEF} = V_{GEF} + V_e$$

De acuerdo con lo señalado más arriba, la variabilidad ambiental puede descomponerse en variabilidad ambiental entre parcelas (V_e) y variabilidad ambiental dentro de parcelas o unidades experimentales (V_{ew}). Este último término es la variabilidad ambiental que contribuye a la variabilidad total entre plantas que comparten una misma unidad experimental. Asumiendo que se esté evaluando una muestra poblacional heterogénea (que todas las plantas poseen diferente constitución genotípica) y que hay independencia de los valores genotípicos respecto a los desvíos ambientales, la variabilidad fenotípica total dentro de parcelas (V_{PW}) resulta de la suma de la variabilidad genotípica dentro de parcelas (V_{GDF}) y la variabilidad ambiental entre plantas dentro de parcelas (V_{ew}) según la siguiente expresión:

$$V_{PW} = V_{GDF} + V_{ew}$$

El tratamiento de la variabilidad dentro de la parcela será diferente en función de si la variable para evaluar resulta de i) observaciones tomadas en una submuestra de individuos dentro de las parcelas, ii) del promedio de las observaciones tomadas en todos los individuos de la parcela, o iii) de la suma de los valores individuales de las plantas de cada parcela. Vencovsky (1980)

describe la relación entre las fuentes de variación y las esperanzas de los cuadrados medios respectivos:

- i. Submuestreo de n plantas individuales en un experimento con r repeticiones (Tabla 5.1):

Fuentes de variación	Cuadrados medios	Esperanza de cuadrados medios
Repeticiones		
Genotipos	CM_g	$\sigma_{we}^2 + \sigma_{wG}^2 + n\sigma_e^2 + rn\sigma_G^2$
Error experimental	CM_e	$\sigma^2 = \sigma_{we}^2 + \sigma_{wG}^2 + n\sigma_e^2$
Dentro de parcelas	CM_d	$\sigma_w^2 = \sigma_{we}^2 + \sigma_{wG}^2$

Tabla 5.1.

Claves del análisis de variancia (ANDeVA) cuando se muestrean individuos dentro de parcelas en experimentos con repeticiones.

Por lo tanto la variabilidad fenotípica (σ_p^2) entre plantas individuales dentro de cada repetición será:

$$\sigma_p^2 = \sigma_{we}^2 + \sigma_{wG}^2 + \sigma_e^2 + \sigma_G^2 = \sigma_w^2 + \sigma_e^2 + \sigma_G^2,$$

donde:

σ_{we}^2 es la variancia ambiental dentro de parcelas,

σ_{wG}^2 es la variancia genotípica dentro de parcelas,

σ_w^2 es la variancia fenotípica dentro de parcelas, cuyo estimador es V_{PW} ,

σ_e^2 es la variancia del error experimental entre parcelas y

σ_G^2 es la variancia genotípica entre familias

- ii. El carácter se mide sobre el total de la parcela (Tabla 5.2):

Fuentes de variación	Cuadrados medios	Esperanza de cuadrados medios
Repeticiones		
Genotipos	CM_g	$\sigma^2 + r\sigma_e^2$
Error experimental	CM_e	$\sigma^2 = n\sigma_w^2 + n^2\sigma_e^2$

Tabla 5.2.

Claves del análisis de variancia (ANDeVA) cuando se trabaja sobre el total a través de n plantas por parcela y r repeticiones

La esperanza del cuadrado medio del error experimental (σ^2) cuando trabajamos con totales de parcelas será:

$$\sigma^2 = n\sigma_w^2 + n^2\sigma_e^2 = n(\sigma_{we}^2 + \sigma_{wG}^2) + n^2\sigma_e^2$$

iii. Cuando se trabaja sobre promedios por parcela (Tabla 5.3):

Fuentes de variación	Cuadrados medios	Esperanza de cuadrados medios
Repeticiones		
Genotipos	CM _g	$\sigma^2 + r\sigma_G^2 = \sigma_w^2/n + \sigma_e^2 + r\sigma_G^2$
Error experimental	CM _e	$\sigma^2 = \sigma_w^2/n + \sigma_e^2$

Tabla 5.3.

Claves del análisis de variancia (ANDeVA) cuando se trabaja con el promedio sobre n individuos dentro de parcelas.

La esperanza del cuadrado medio del error experimental (σ^2) cuando trabajamos con los promedios de observaciones sobre n plantas por parcela será:

$$\sigma^2 = (\sigma_{we}^2 + \sigma_{wG}^2)/n + \sigma_e^2$$

5.2. Variabilidad genotípica

La variabilidad existente para un carácter de interés en una población que responde a causas genéticas depende del número de loci, de la diversidad alélica y de los modos de acción génica. En capítulos previos revisamos cómo esos dos primeros factores determinan la existencia de individuos genéticamente diferentes y en este capítulo nos ocuparemos de cuantificar y describir esta variabilidad a nivel poblacional.

5.2.1. Cuantificación de la variancia genotípica de poblaciones experimentales derivadas del cruzamiento entre progenitores homocigotas

5.2.1.1. Variancia genotípica en poblaciones F₂

Recordemos que por definición (ver Anexo) la variancia es la suma de los desvíos al cuadrado entre los valores individuales y su media general. Por lo tanto la variancia genotípica será la suma de cuadrados de los desvíos entre los valores genotípicos de cada clase genotípica, ponderados por su frecuencia y la media poblacional. La Tabla 5.4 presenta la manera de calcular la variancia genotípica de una población F₂, derivada del cruzamiento entre dos líneas homocigotas contrastantes para un locus con dos alelos.



Clase genotípica	Frecuencia genotípica (1)	Cuadrado de los desvíos respecto de la media genotípica (2)	(1)x(2)
L ₁ L ₁	¼	(a - ½ d) ² = a ² - ad + ¼ d ²	¼(a ² - ad + ¼ d ²)
L ₁ L ₂	½	(d - ½d) ² = ¼ d ²	½ ¼ d ²
L ₂ L ₂	¼	(-a - ½ d) ² = a ² + ad + ¼ d ²	¼(a ² + ad + ¼ d ²)
Variancia F₂		½ a² + ¼ d²	

Tabla 5.4.

Variancia genotípica entre individuos F₂ derivados por autofecundación de individuos F₁ del cruzamiento entre líneas homocigotas y polimórficas entre sí para el locus L.

En este caso la variancia de la filial F₂ se explica por la contribución de los valores aditivos y de los desvíos de dominancia:

$$V_{F_2} = (1/2)a^2 + (1/4)d^2, \quad [5.5]$$

Los términos del miembro de la derecha de la igualdad corresponden a la variancia aditiva y a la variancia de dominancia, respectivamente. Esta expresión puede simplificarse cuando no existan desvíos de dominancia [5.5.1], o que exista dominancia completa [5.5.2]:

$$V_{F_2} = (1/2)a^2, \quad [5.5.1]$$

$$V_{F_2} = (3/4)a^2, \quad [5.5.2]$$

5.2.1.2. Variancia genotípica en poblaciones derivadas del retrocruzamiento de la F₁ a cada progenitor

Cuando la filial F₁ se retrocruza al padre portador de los alelos favorables o al padre con los alelos desfavorables, pueden deducirse las variancias correspondientes (Tabla 5.5):

Padre recurrente	Clase genotípica	Frecuencia genotípica (1)	Cuadrado de los desvíos de la media genotípica (2)	(1)x(2)
P ₁	L ₁ L ₁	½	¼ (a ² - 2ad + d ²)	1/8 (a ² - 2ad + d ²)
	L ₁ L ₂	½	¼ (a ² - 2ad + d ²)	1/8 (a ² - 2ad + d ²)
Variancia de RC₁P₁				¼ (a² - 2ad + d²)
P ₂	L ₂ L ₂	½	¼ (a ² + 2ad + d ²)	1/8 (a ² + 2ad + d ²)
	L ₁ L ₂	½	¼ (a ² + 2ad + d ²)	1/8 (a ² + 2ad + d ²)
Variancia de RC₁P₂				¼ (a² + 2ad + d²)

Tabla 5.5.

Variancias genotípicas de la generación de primer retrocruzamiento a los padres homocigotas para el alelo favorable (P₁) y para el alelo desfavorable (P₂).

$$V_{RC1.P1} = (1/4)(a^2 - 2ad + d^2), \quad [5.6.1]$$

$$V_{RC1.P2} = (1/4)(a^2 - 2ad + d^2), \quad [5.6.2]$$

5.2.1.3. Variancia genotípica en una población endocriada hasta homocigosis (de líneas recombinantes o haploides duplicados)

En esta circunstancia (Tabla 5.6) la variancia simplemente se reduce a:

$$V_{F_n} = a^2, \text{ donde } F_n \text{ se refiere a la filial en la que el coeficiente de endocría es } F=1.$$

Clase genotípica	Frecuencia genotípica (1)	Desvíos al cuadrado respecto de la media (2)	(1)x(2)
L ₁ L ₁	½	a ²	½ (a) ²
L ₂ L ₂	½	(-a) ²	½ (-a) ²
Variancia F_n			a²

Tabla 5.6.

Variancia genotípica entre líneas recombinantes o generadas por duplicaciones de haploides (F_n) a partir de individuos F₂ derivados por autofecundación o apareamiento de individuos F₁ del cruzamiento entre líneas homocigóticas y polimórficas entre sí para el locus L.

5.2.2. Variancia genotípica y sus componentes causales en una población no endocriada y bajo equilibrio de Hardy-Weinberg

Recordemos las expresiones de los valores genotípicos y de sus dos componentes, aditivo y de dominancia, expresados en función de los efectos medios de sustitución de un gen y centrados en la media poblacional para los genotipos posibles en una población bajo equilibrio de Hardy-Weinberg (Tabla 3.13).

Los valores genotípicos (G) de los individuos de una población, en ausencia de epistasis y considerando solo un locus, tienen como componentes a los valores aditivos (A) y de dominancia (D). Al describir la variancia genotípica en función de sus componentes causales debemos notar que estamos ante una función lineal del tipo:

$$G = A + D$$

En consecuencia $V_G = V_{(A+D)} = V_A + V_D + 2Cov_{AD}$

donde:

V_G es la variancia genotípica;

V_A es la variancia aditiva;

V_D es la variancia de dominancia; y

Cov_{AD} es la covariancia entre los valores aditivos y de dominancia.

5.2.2.1. Variancia aditiva y variancia de dominancia

En el modelo

$$V_G = V_A + V_D + 2Cov_{AD}$$

y en ausencia de endocría, los valores aditivos (A) y los desvíos de dominancia (D) son independientes, por lo cual la covariancia entre los valores aditivos y los de dominancia es igual a cero.

Consecuentemente:

$$V_G = V_{(A+D)} = V_A + V_D$$

Por una parte, las variancias se calculan como sumatoria de los cuadrados de los respectivos desvíos entre los valores y la media. Como los valores ya están centrados en la media poblacional, los desvíos con respecto a la media son directamente los valores genotípicos, aditivos y de dominancia. Por otra parte, la contribución relativa de cada clase genotípica a la variancia poblacional dependerá de la cantidad de individuos en cada una, por lo que los valores deben ponderarse por las frecuencias genotípicas esperadas bajo las condiciones de equilibrio de Hardy-Weinberg. Estas variancias se presentan en la Tabla 5.7.

Clases Genotípicas	Frec. x (Val. Genotíp.) ²	Frec. x (Val. Aditivos) ²	Frec. x (Desvíos de Dominancia) ²
--	$(q^2)[-2p(a + pd)]^2$	$(q^2)(-2pa)^2$	$(q^2)(-2p^2d)^2$
+-	$(2pq)[(q-p)a + 2pqd]^2$	$(2pq)[(q-p)a]^2$	$(2pq)(2pqd)^2$
++	$(p^2)[2q(a - dq)]^2$	$(p^2)(2qa)^2$	$(p^2)(-2q^2d)^2$
	Genotípica	Aditiva o Genética	De Dominancia
Variancias	$2pq a^2 + (2pqd)^2 = 2pq[a+d(q-p)]^2 + (2pqd)^2$	$2pq a^2 = 2pq[a + d(q-p)]^2$	$(2pqd)^2$

Tabla 5.7.

Contribución de las clases genotípicas (un locus, dos alelos) en una población en equilibrio de Hardy-Weinberg a la variancia genotípica y a las variancias aditiva y de dominancia.

A modo de comprobación, nótese que la suma de la variancia aditiva y la de dominancia iguala a la variancia genotípica, por cuanto en el modelo estamos considerando un locus y por lo tanto no hay posibilidad de interacción entre loci o epistasis. A la variancia aditiva normalmente se la denomina variancia genética y no es otra cosa que la variancia de los valores aditivos (valores reproductivos o valores de mejora) que existe entre los individuos que componen una población.

En casos especiales las expresiones de la variancia genotípica (V_G) y de la variancia aditiva (V_A) pueden simplificarse del modo siguiente:

Si el desvío de dominancia es nulo ($d=0$):

$$V_G = V_A = 2pq\alpha^2 = 2pqa^2$$

Cuando la dominancia es completa ($d=a$):

$$V_G = V_A + V_D = 8p^3qa^2 + 4p^2q^2a^2$$

Cualesquiera sean los niveles de dominancia, si $p=q=0,5$:

$$V_G = V_A + V_D = (1/2)a^2 + (1/4)d^2$$

Cuando $p=q=0,5$ y $d=0$:

$$V_G = V_A = (1/2)a^2$$

Dados $p = q = 0,5$ y $d = a$:

$$V_G = V_A + V_D = (1/2)a^2 + (1/4)a^2 = (3/4)a^2$$

Obsérvese que la variancia genotípica y sus componentes en una población en equilibrio de Hardy-Weinberg con frecuencias génicas $p=q=0,5$ es igual a la variancia de una población F_2 derivada del cruzamiento de dos progenitores homocigotas. La variancia genética depende únicamente del efecto promedio de sustitución de un gen. A menos que el desvío de dominancia (d) sea cero o las frecuencias génicas sean intermedias, dependerá tanto del promedio de los homocigotas (a) como de los desvíos de dominancia (d).

Como se desprende de los párrafos anteriores y de la Tabla 5.7, la variancia genotípica y sus componentes de variancia genética y de dominancia para un determinado carácter dependen de las frecuencias génicas en la población y por lo tanto son parámetros poblacionales. Cualquier proceso sistemático que altere las frecuencias génicas tendrá como consecuencia la modificación de la variancia genotípica y la magnitud y relación entre sus componentes aditivo y de dominancia. Veremos más adelante que los

procesos dispersivos generados por pequeños tamaños poblacionales o sistemas de endogamia también afectan a la variancia genotípica poblacional y sus componentes.

5.2.2.2. *Variancia epistática*

La variabilidad en los efectos de interacción entre loci da lugar a la existencia de la variancia epistática. La epistasis estadística es una propiedad de la población y por lo tanto se modifica con las frecuencias génicas. Existirá variancia de interacción epistática cuando los efectos de sustitución de un gen o los desvíos de dominancia en un locus se modifican según cual sea el genotipo en otro locus.

La variancia de interacción epistática (V_I) puede subdividirse en diferentes componentes teniendo en cuenta distintos criterios: el número de loci involucrados, los valores o tipos de acción génica involucradas, y ambos criterios a la vez. Así podemos hablar de variancia epistática de interacción entre dos factores (dos loci), tres factores (o tres loci), n factores. Dependiendo de los efectos involucrados podremos distinguir las variancias de interacción entre los valores de mejora, entre los desvíos de dominancia y de interacción entre los valores de mejora y los desvíos de dominancia:

$$V_G = V_A + V_D + V_I = V_A + V_D + V_{AA} + V_{AD} + V_{DD} + V_{AAA} + V_{AAD} + V_{ADA} + V_{DAA} + V_{ADD} + V_{DDA} + V_{DAD} + V_{DDD}$$

donde:

los subíndices A y D hacen referencia a valores aditivos y desvíos de dominancia cuya variabilidad entre individuos contribuye a las variancias correspondientes,

($V_{AA} + V_{AD} + V_{DD}$) son los componentes de la variancia de interacción epistática de dos factores, aditiva x aditiva, aditiva x dominante, dominante x dominante y

($V_{AAA} + V_{AAD} + V_{ADA} + V_{DAA} + V_{ADD} + V_{DDA} + V_{DAD} + V_{DDD}$) son los componentes de la variancia de interacción epistática de tres factores.

Si nos situamos en el caso de dos loci con dos alelos cada uno, el análisis de la epistasis guarda analogía con el análisis de variancia de un diseño estadístico de dos factores (loci) a tres niveles (0, 1 y 2 alelos favorables por locus) (Tabla 5.8).

Fuentes de variación	Modelo genético	
Efectos principales		
Lineal	Aditividad	A
Cuadrático	Dominancia	D
Interacciones		
Epistasis		
Lineal x Lineal	Aditiva x Aditiva	I _{AA}
Lineal x Cuadrático	Aditiva x Dominante	I _{AD}
Cuadrático x Cuadrático	Dominante x Dominante	I _{DD}

Tabla 5.8. Análisis factorial de la interacción epistática entre dos loci (A y B) con dos alelos por locus.

5.3. La variancia genotípica y sus componentes causales en una población endocriada

La existencia de endocria en la población provoca que los efectos aditivos y de dominancia no sean independientes generando una Cov_{AD} diferente de cero. En consecuencia la variancia genotípica, aun en ausencia de epistasis, no coincidirá con la suma de las variancias aditiva y de dominancia, excepto en casos especiales.

La expresión matemática de la variancia aditiva en una población endocriada es similar a la de una población no endocriada, pero requiere una corrección por el coeficiente de endogamia (F). Conforme aumente el coeficiente de endogamia, aumentará también la variancia aditiva (V_A') según la expresión:

$$V_A' = 2pq(1 + F)^F \alpha^2 = 2pq(1 + F) |a + [(1 - F)/(1 + F)]d(q - p)|^2$$

Por el contrario, dependiendo de las frecuencias génicas, la V_D' disminuye proporcionalmente con el coeficiente de endogamia:

$$V_D' = 4pq [(1 - F)/(1 + F)](p + Fq)(q + Fp) d^2$$

Si las frecuencias génicas poblacionales son $p=q=0,5$ y el coeficiente de endogamia es $F=1$, la variancia de dominancia se hace nula y la variancia aditiva se duplica con respecto a la de la misma población en ausencia de endocria. La Tabla 5.9 presenta las variancias genotípicas y sus componentes para cualquier nivel de endocria.

Clases genotíp.	(Frec. x Val. Genotíp.) ²	(Frec. x Val. Aditivos) ²	(Frec. x Desvíos de Dominancia) ²
--	$(q^2 + pqF)[-2p(Fa + pd)]^2$	$(q^2 + pqF)(-2pFa)^2$	$(q^2 + pqF)(-2p^2d)^2$
+-	$(2pq)(1-F)[(q-p)Fa + 2pqd]^2$	$2pq(1-F)[(q-p)Fa]^2$	$(2pq)(1-F)(2pqd)^2$
++	$(p^2 + pqF)[2q(Fa - dq)]^2$	$(p^2 + pqF)(2qFa)^2$	$(p^2 + pqF)(-2q^2d)^2$
	Genotípica	Aditiva	De dominancia
Variaciones	$2pq(1+F)[a+[(1-F)/(1+F)]d(q-p)]^2 + 4pq[(1-F)/(1+F)](p+Fq)(q+Fd)d^2$	$2pq(1+F)Fa^2 = 2pq(1+F)[a+[(1-F)/(1+F)]d(q-p)]^2$	$4pq[(1-F)/(1+F)](p+qF)(q+pF)d^2$

Tabla 5.9. Contribución a la variancia genotípica y a la variancia aditiva y de dominancia de las clases genotípicas (un locus, dos alelos) en una población en equilibrio de Hardy-Weinberg con endogamia previa.

Puede comprobarse que si $F=0$ se llega a las mismas expresiones de variancias genotípicas y sus variancias componentes causales presentadas en la Tabla 5.7.

En ausencia de dominancia, las expresiones se simplifican, de modo que:

$$V_A' = 2pq(1 + F)a^2 \text{ y } V_D' = 0$$

En el caso de existir dominancia solo si $p = q = 0.5$ es posible expresar las relaciones entre V_A' y V_A , y entre V_D' y V_D (Hallauer *et al.*, 1988):

$$V_A' = (1 + F)V_A \text{ y } V_D' = (1 - F^2)V_D$$

Nótese que en caso de que $F=0$, las expresiones igualan a las indicadas más arriba para poblaciones con frecuencias génicas intermedias:

$$V_A = (1/2)a^2$$

y

$$V_D = (1/4)d^2$$

Las variancias genotípicas, y los componentes de variancia aditiva (V_A') y de dominancia (V_D') se van modificando según el coeficiente de endogamia y esta modificación depende de las frecuencias génicas para el alelo favorable (Figura 5.1).

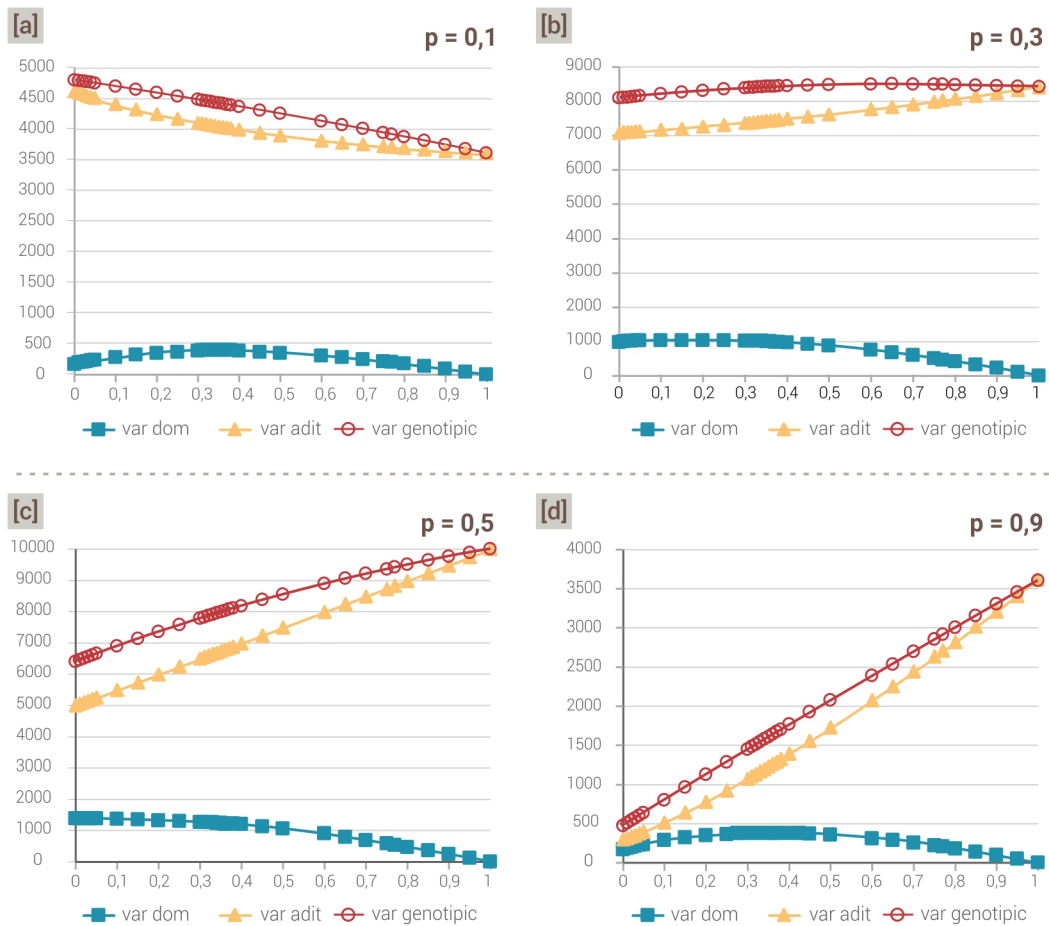


Figura 5.1. Variancias aditiva, de dominancia y genotípica para distintos coeficientes de endogamia y diferentes frecuencias génicas del alelo favorable: [a], $p=0,1$; [b], $p=0,25$; [c], $p=0,5$; [d], $p=0,9$. Se considera que $a=100$ y $d=75$.

A menos que la frecuencia del alelo favorable sea baja, el incremento de la endogamia provoca incrementos en la variancia aditiva y en la variancia genotípica. Si la frecuencia del alelo favorable es baja, existe una primera fase en la que la variancia aditiva disminuye. Cuando la frecuencia génica es $p=q=0,5$, el aumento de la endogamia provoca incrementos lineales de la variancia aditiva. La marcha de la variancia de dominancia con el aumento de endogamia sigue una curva exponencial hasta hacerse igual a cero cuando $F=1$ (Figuras 5.1 y 5.2).



[a] **Variación Aditiva**

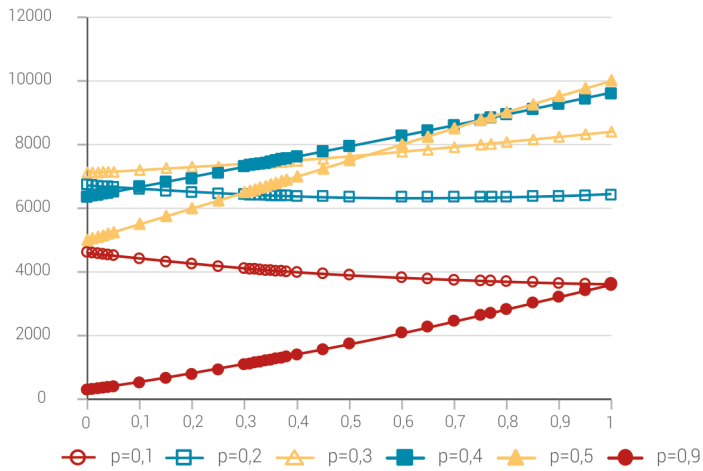
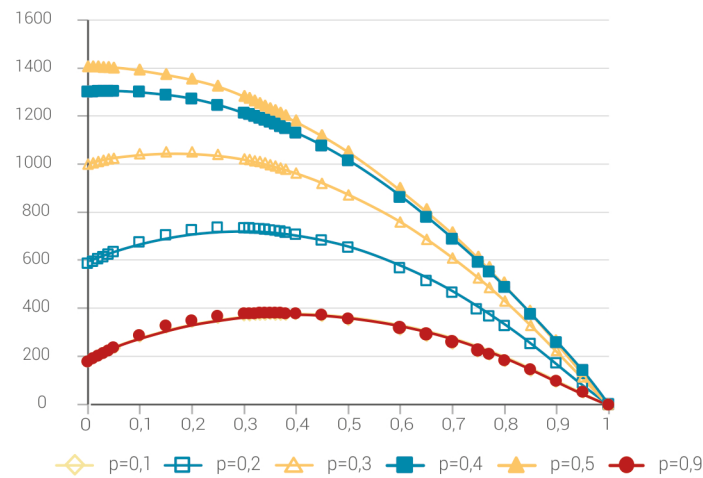
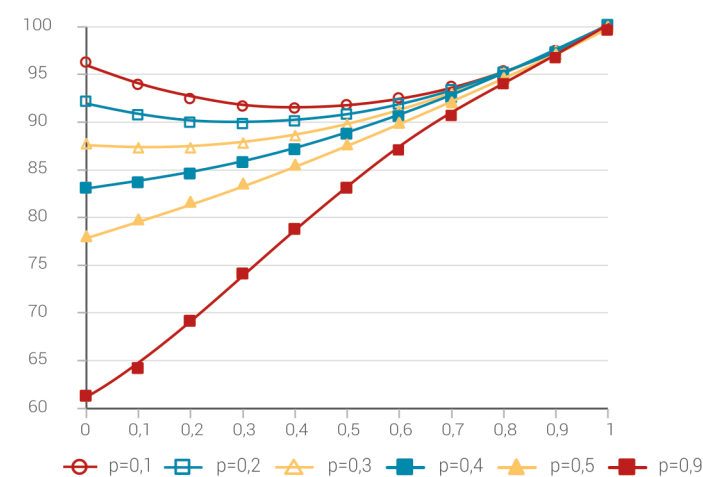


Figura 5.2. Variación aditiva [a], de dominancia [b] y relación entre variación aditiva y genotípica [c] en función del coeficiente de endogamia para diferentes frecuencias génicas del alelo favorable. Se asume $a=100$ y $d=75$.

[b] **Variación de Dominancia**



[c] **Var. Aditiva/Var. Genotípica x 100**



Cuando la endogamia es máxima, la mayor variancia aditiva ocurre a frecuencias alélicas intermedias. Si la frecuencia del alelo favorable se acerca a los niveles de pérdida o fijación, la variancia será menor (Figura 5.2). Cuando la frecuencia del alelo favorable es baja, la variancia aditiva se mantiene relativamente estable con el incremento de la endogamia. La relación entre las variancias aditiva y dominante se modifica drásticamente en función del coeficiente de endogamia y de las frecuencias génicas. A mayor endogamia, habrá preponderancia de la variancia aditiva respecto de la de dominancia. Estas proporciones serán más favorables desde el punto de vista de la selección si los alelos favorables en la población se encuentran en frecuencias intermedias a altas.

En el caso de que la acción génica que controle el carácter sea aditiva, conforme se incremente el coeficiente de endogamia de la población por aplicación de algún sistema de endocria o por deriva genética, aumentará también la variancia aditiva, a menos que la aditividad coexista con interacciones intralocus e interloci. En presencia de interacciones aditiva x aditiva, la variancia aditiva puede ir aumentando hasta alcanzar un máximo cuando $F \approx 0,5$ y luego decrecer hasta anularse con $F=1$. En todo el rango de F , la variancia poblacional resulta exponencialmente creciente. El incremento inicial de la variancia aditiva puede ocurrir por modificaciones en el efecto medio de los alelos a medida que la endocria va modificando el fondo genético donde ellos se expresan (Goodnight, 1999).

La endocria provoca cambios en la variancia genética, incrementando la variancia entre líneas y disminuyéndola dentro de cada una de ellas con niveles crecientes de F . Estos cambios son independientes de las frecuencias génicas en ausencia de dominancia y epistasis. Diversos estudios han comprobado que frecuentemente las líneas endocriadas son más sensibles o vulnerables a los efectos del ambiente, por lo que puede suceder que también ocurra un incremento en la variancia ambiental.

Cuando existe dominancia no puede deducirse un modelo general para explicar los cambios en la distribución de la variancia genética, aunque pueden calcularse a condición de conocer las frecuencias génicas (Falconer y Mackay, 2001).

5.4. Elección de probadores y variancia entre cruzamientos de prueba. Consideraciones teóricas e implicancias prácticas

En los programas de mejoramiento orientados al desarrollo de híbridos entre líneas endocriadas es necesario disponer de un esquema de evaluación de esas líneas en cruzamientos de prueba con un probador o *tester* que permita evaluar comparativamente la habilidad combinatoria y discriminar correctamente el mérito genético relativo de esas líneas. La elección del probador, entonces, constituye una decisión crítica por cuanto afecta una de las etapas claves del desarrollo y selección de líneas progenitoras de híbridos. Por razones análogas, también es crucial tener en cuenta las características del probador en programas de selección recurrente.

Rawlings y Thompson (1962) aportaron las consideraciones teóricas sobre la elección de líneas probadoras para la evaluación de líneas endocriadas. Considerando el objetivo de diferenciar apropiadamente el mérito genético de un grupo de genotipos, el mejor probador (*tester*) será, por definición, el que maximice la variabilidad genética entre los cruzamientos de prueba. Los autores asumieron un modelo con niveles variables de dominancia y con ausencia de epistasis y concluyeron que la variancia genética entre cruzamientos de prueba será:

$${}^{\text{CP}}\sigma_g^2 = (1/2)(pq)[1 + F][a + (1 - 2r)d]^2,$$

donde:

p y q son las frecuencias génicas del alelo favorable y del desfavorable, respectivamente, en la población de cría de la que se están derivando líneas endocriadas,

r es la frecuencia del alelo favorable en el probador y

F es el coeficiente de endocría de la población de líneas.

Por una parte, para una misma condición de dominancia y para una frecuencia de probador determinada, la variancia entre cruzamientos de prueba se hace mayor cuando la frecuencia alélica del gen favorable en la población que se está evaluando o seleccionando se encuentra en niveles intermedios (0,5) (Figura 5.3).

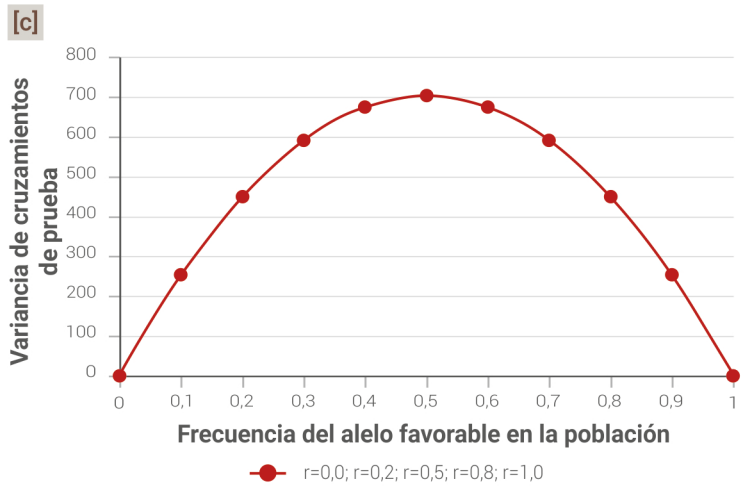
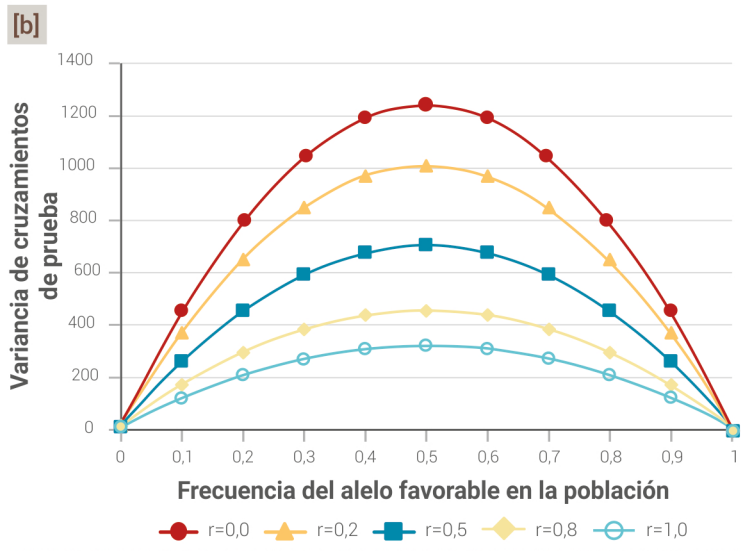
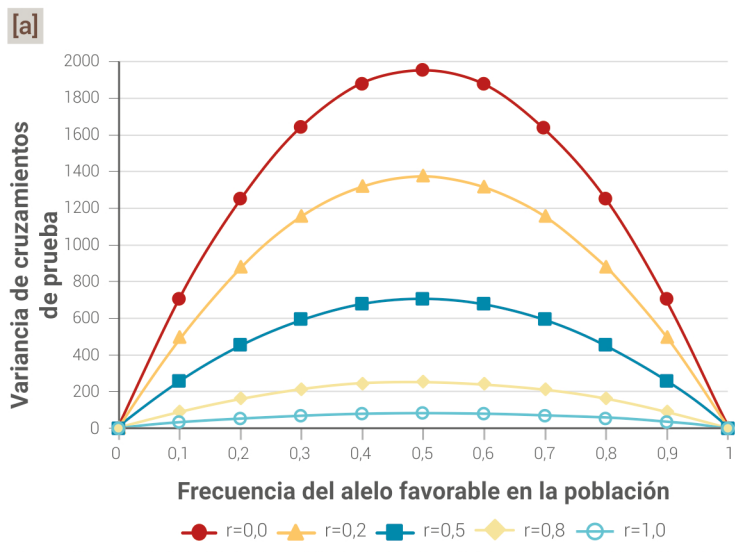


Figura 5.3. Variación de cruzamientos de prueba, considerando tres valores de desvíos de dominancia (a: $d=50$; b: $d=25$ y c: $d=0$), un valor aditivo $a=75$, en función de la frecuencia del alelo favorable en la población de referencia, y para distintas frecuencias del alelo favorable en el probador (r).

Por otra parte, para cualquier frecuencia del alelo favorable en la población, la variancia entre cruzamientos de prueba será mayor cuanto menor sea la frecuencia del alelo favorable en el probador (r), a menos que no exista dominancia (Figura 5.4).

La variancia de cruzamientos de prueba no solamente es función de las características del probador (sus frecuencias génicas), sino que también depende de las frecuencias en el material que se somete a evaluación.

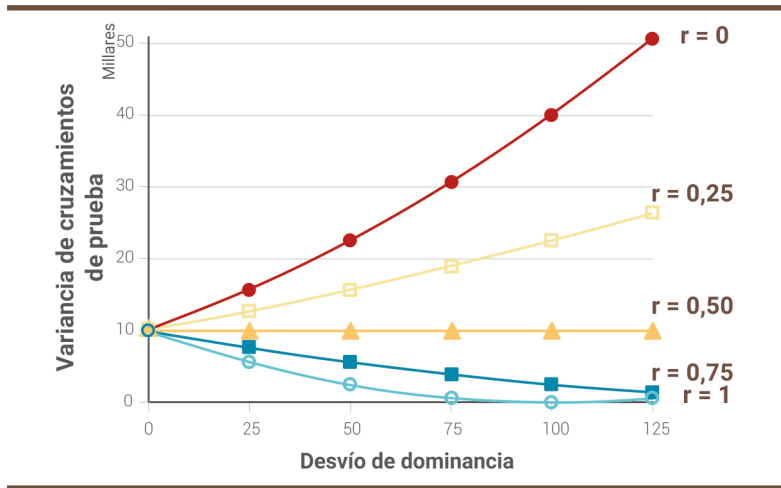


Figura 5.4. Efecto de diferentes frecuencias del alelo favorable en el probador sobre la variancia de cruzamientos de prueba, considerando un locus $a=100$ y diferentes desvíos de dominancia.

Considerando distintos niveles de dominancia, en la Figura 5.5 se grafica la variancia de cruzamientos de prueba de dos probadores homocigotas, con frecuencia 0 y 1 para los alelos favorables al carácter y diferentes frecuencias de alelos favorables en la población bajo evaluación en caso de emplearla como probadora.

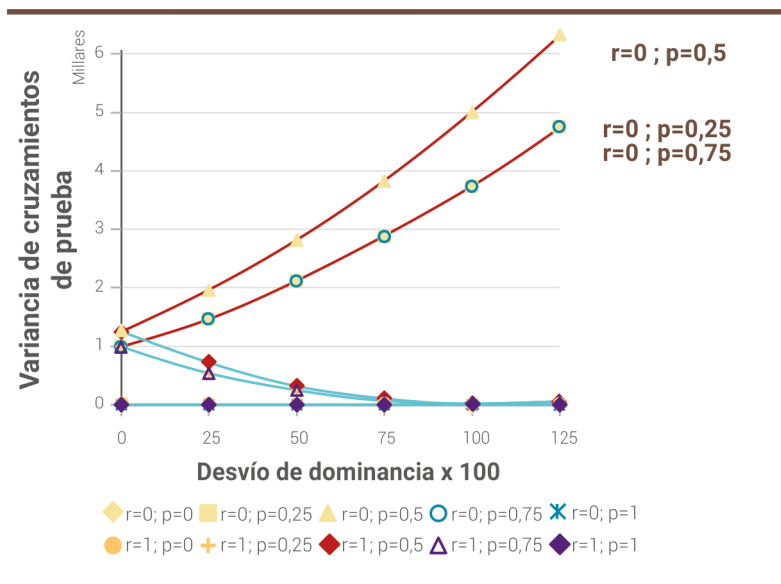


Figura 5.5. Variancia entre cruzamientos de prueba de poblaciones con diferentes frecuencias de alelos favorables (p) utilizando dos probadores alternativos (frecuencias para los alelos favorables $r=0$ y $r=1$) para diferentes niveles de dominancia del carácter. Se asume $a=100$.

Pueden extraerse tres conclusiones básicas a partir del análisis e interpretación de las Figuras 5.3, 5.4 y 5.5. En primer lugar, que a igualdad de otras condiciones, las poblaciones de cría originadas a partir de la autofecundación de un cruzamiento simple son las que permitirán manifestar mayor variancia entre cruzamientos de prueba y probablemente mayor heredabilidad para seleccionar entre líneas. En segundo término, el mejor probador será, claramente, el que posee la menor frecuencia de alelos favorables, cualquiera sea la frecuencia del alelo favorable en la población de referencia. Esto es cierto, si y solo si existe dominancia. En ausencia de dominancia ($d=0$) no importa la constitución alélica del probador porque no afecta la magnitud de la variancia entre cruzamientos de prueba. En tercer término, cuando $d \neq 0$ la variancia de cruzamientos de prueba será mayor cuanto mayor sea la diferencia de frecuencias génicas entre la población bajo prueba y el probador. A los fines prácticos, eso explica el interés de clasificar el germoplasma de un programa de mejoramiento en grupos heteróticos caracterizados por poseer diferencias génicas en los loci que contribuyen al rendimiento. De tal forma, las líneas que se derivan de una población de referencia, se evalúan con un probador de un grupo heterótico opuesto. Utilizando esta estrategia se espera que la diferencia entre las frecuencias génicas entre probador y población probada ($p - r$) sea importante.

En ausencia de dominancia la variancia entre los cruzamientos de prueba no se ve afectada por la constitución genética del probador y tampoco cuando el probador tiene frecuencia génica del alelo favorable igual a 0,5, para cualquier nivel de dominancia. Si la dominancia está en el rango de parcial a completa y la frecuencia génica del alelo favorable en el probador es mayor a 0,5, la variancia resultará decreciente y si existiera sobredominancia cuanto más importante sea la sobredominancia la variancia de cruzamientos de prueba será creciente. En cualquier caso y para todo el rango de niveles de dominancia, se esperan mayores variancias de los cruzamientos de prueba cuando el probador posea frecuencias génicas del alelo favorable inferiores a 0,5.

De la Figura 5.4 y la expresión indicada más arriba, el mejor probador sería aquel que carece de alelos favorables en la mayoría de los loci que controlan el carácter. Allison y Curnow (1966) analizaron el efecto de la constitución del probador en el mejoramiento de poblaciones mediante selección recurrente de medios hermanos basada en cruzamientos de prueba y arribaron a conclusiones similares a las de Rawlings y Thompson (1962). Si bien la existencia de epistasis puede afectar la variancia entre cruzamientos de prueba, el sesgo que ella provocaría respecto de lo que indica el modelo no sería importante (Rawling y Thompson, 1962).

5.5. Variancias genotípicas entre híbridos F₁

Asumiendo que se dispone de un número determinado de líneas endocriadas y se evaluarán la totalidad de los híbridos de dos, tres y cuatro líneas posibles de generar con las diferentes combinaciones de tales líneas, las variancias genotípicas dentro de cada tipo de híbridos en términos relativos a los híbridos simples se presentan en la Tabla 5.10.

Tipo de híbrido	Variancia				
	Aditiva	Dominancia	Epistática		
			AA	AD	DD
Simple (2x)	1	1	1	1	1
Tres líneas (3x)	3/4	1/2	9/16	3/8	1/4
Dobles (4x)	1/2	1/4	1/4	1/8	1/16

Tabla 5.10.

Componentes de la variancia genotípica para diferentes tipos de híbridos entre líneas homocigotas.

Consecuentemente, un experimento destinado a evaluar y seleccionar (elegir) los mejores híbridos presentará mayor variabilidad si se trata de híbridos simples, que si se trata de híbridos de tres y cuatro líneas, mejorando así las posibilidades de discriminar entre híbridos para identificar y seleccionar los mejores y hacerlo con mayor confiabilidad.

5.6. Variancia de cruzamientos interpoblacionales

Dadas dos poblaciones, P₁ y P₂, con frecuencias génicas (p, q) y (r, s), respectivamente, podemos definir las variancias aditivas intrapoblacionales como:

$$\sigma_{A1}^2 = 2pq\alpha_1^2$$

$$\sigma_{A2}^2 = 2rs\alpha_2^2$$

donde:

α_1 es el efecto de sustitución del gen en la población P₁,

α_2 es el efecto de sustitución del gen en la población P₂,

y a los homólogos de variancias genéticas o aditivas interpoblacionales como (Miranda-Filho, 1982):

$$\sigma_{A12}^2 = 2pq\alpha_2^2$$

$$\sigma_{A21}^2 = 2rs\alpha_1^2$$

Asimismo las covariancias interpopulacionales aditivas serán:

$$\sigma_{A1A2} = 2pq\alpha_1\alpha_2$$

$$\sigma_{A2A1} = 2rs\alpha_1\alpha_2$$

Souza Jr. (1992, 1999) demostró que las variancias genéticas interpopulacionales pueden expresarse como:

$$\sigma_{A12}^2 = \sigma_{A1}^2 + \sigma_{\tau12}^2 + 4Cov_{A1,\tau12}$$

$$\sigma_{A21}^2 = \sigma_{A2}^2 + \sigma_{\tau21}^2 + 4Cov_{A2,\tau21}$$

donde $\sigma_{\tau ij}^2$ es la variancia de los desvíos entre los efectos aditivos de la población *i*-ésima y los efectos aditivos interpopulacionales del cruzamiento entre la *i*-ésima y *j*-ésima población.

Si las frecuencias génicas en las poblaciones P_1 y P_2 son (p, q) y (r, s), respectivamente,

$$\sigma_{\tau12}^2 = 8pq(p-r)^2d^2$$

$$\sigma_{\tau21}^2 = 8rs(r-p)^2d^2$$

A su vez,

$$\sigma_{A1A2} = Cov_{A1A2} = \sigma_{A1}^2 + 2Cov_{A1,\tau12}$$

$$\sigma_{A2A1} = Cov_{A2A1} = \sigma_{A2}^2 + 2Cov_{A2,\tau21}$$

Por lo tanto:

$$Cov_{A1,\tau12} = (\sigma_{A1A2} - \sigma_{A1}^2)/2$$

$$Cov_{A2,\tau21} = (\sigma_{A2A1} - \sigma_{A2}^2)/2$$

$Cov_{A_i,\tau ij}$ es la covariancia de los efectos aditivos de la población *i*-ésima con los desvíos entre los efectos aditivos intrapoblacionales de la población *i*-ésima con los interpopulacionales del cruzamiento entre la *i*-ésima y *j*-ésima población. Expresando estas covariancias según las frecuencias génicas de las poblaciones P_1 y P_2 :

$$Cov_{A1,\tau12} = 2pq(p-r)\alpha_1d$$

$$Cov_{A2,\tau21} = 2rs(r-p)\alpha_2d$$

De tal manera:

$$\sigma_{A12}^2 = 2pq\alpha_1^2 + 8pq(p-r)^2d^2 + 8pq(p-r)\alpha_1d$$

$$= 2pq[\alpha_1^2 + 4(p-r)^2d^2 + 4(p-r)\alpha_1d]$$

$$= 2pq[\alpha_1 + 2d(p-r)]^2$$

y

$$\sigma_{A21}^2 = 2rs\alpha_2^2 + 8rs(r-p)^2d^2 + 8rs(r-p)\alpha_2d$$

$$= 2rs[\alpha_2^2 + 4(r-p)^2d^2 + 4(r-p)\alpha_2d]$$

$$= 2rs[\alpha_2 + 2d(r-p)]^2$$

5.7. Covariancia entre valores aditivos y de dominancia en poblaciones endocriadas y no endocriadas

En el modelo inicial aplicado a una población sin endocría previa, los efectos aditivos y los de dominancia resultan independientes y por lo tanto su covariancia es igual a cero. Por el contrario, cuando la población es endogámica, la condición de independencia entre aquellos efectos no se cumple. Puede deducirse que cuanto mayor sea la endogamia, mayor será la covariancia entre los valores aditivos y los desvíos de dominancia, a menos que las frecuencias génicas sean intermedias (Tabla 5.11). Esta última situación se trata de un caso especial, en el que a pesar de la existencia de endocría, la covariancia es cero. La propiedad de las poblaciones con frecuencias génicas intermedias en cuanto a sus componentes causales de variancia ha dado lugar a diseños genéticos particulares (ver Capítulo 8).

Población no endocriada					Población endocriada			
Genotipo	Frec. Gen.	A	D	(Frec. x A x D)	Frec. Gen.	A	D	(Frec. x A x D)
--	q ²	-2pa	-2p ² d	q ² (-2pa) (-2p ² d)	q ² + pqF	-2pa	-2p ² d	(q ² +pqF)(-2pa) (-2p ² d)
+-	2pq	(q-p)a	2pqd	2pq(q-p) a(2pqd)	2pq(1+F)	(q-p)a	2pqd	2pq(1+F)(q-p) a(2pqd)
++	p ²	2qa	-2q ² d	p ² (2qa) (-2q ² d)	p ² + pqF	2qa	-2q ² d	(p ² +pqF)(2qa) (-2q ² d)
Cov (A,D)				0	Cov (A,D)			4pq(p-q)Fad

Tabla 5.11.

Covariancia entre los valores aditivos y de dominancia en poblaciones en equilibrio de Hardy-Weinberg, con y sin endogamia. A y D son los valores aditivos y los desvíos de dominancia, respectivamente. Se asumen frecuencias génicas p y q para un locus con dos alelos.

5.8. Desequilibrio gamético

Recordemos que el desequilibrio gamético ocurre cuando las frecuencias de algunas gametas para dos o más loci tomados conjuntamente son diferentes a las esperadas en función de las frecuencias gaméticas teóricas (ver Capítulo 2).

La existencia de ligamiento genético implica la falta de independencia entre los efectos genotípicos de los loci involucrados. Esta falta de independencia suma a la contribución de ambos loci a la variancia genotípica total, un término de covariancia entre estos. Si un carácter está controlado por n loci, la variancia genotípica total será:

$$V_G = V_{G(1,2,3...n)} = V_{G1} + V_{G2} + V_{G3} + \dots + V_{Gn} + 2[\text{Cov}_{(G1,G2)} + \text{Cov}_{(G1,G3)} + \dots + \text{Cov}_{(Gn-1,Gn)}] \quad [5.5]$$

Cada uno de los términos de covariancia será distintos de cero si los loci involucrados estuvieran ligados, y serán iguales a cero si hubiera independencia.

En ausencia de epistasis, cada uno de los términos de variancia, V_{Gi} , corresponde a la suma de las contribuciones a la variancia genotípica de la variancia aditiva y la variancia de dominancia. De manera análoga, cada término de covariancia entre pares de genes será (Bernardo, 2010):

$$\text{Cov}_{(G1,G2)} = \text{Cov}_{(\alpha1+\delta1; \alpha2+\delta2)} = 2\text{Cov}_{(\alpha1,\alpha2)} + \text{Cov}_{(\delta1,\delta2)}$$

donde:

α_1 y α_2 hacen referencia al valor de mejora debida a los locus 1 y 2, respectivamente,

δ_1 y δ_2 son los desvíos de dominancia de los locus 1 y 2, respectivamente.

Por lo tanto, reemplazando en [5.5]:

$$V_G = V_{G(1,2,3...n)} = V_{A1} + V_{A2} + V_{A3} + \dots + V_{An} + V_{D1} + V_{D2} + V_{D3} + \dots + V_{Dn} + 4[\text{Cov}_{(\alpha1,\alpha2)} + \text{Cov}_{(\alpha1,\alpha3)} + \dots + \text{Cov}_{(\alpha n-1,\alpha n)}] + 2[\text{Cov}_{(\delta1,\delta2)} + \text{Cov}_{(\delta1,\delta3)} + \dots + \text{Cov}_{(\delta n-1,\delta n)}]$$

En maíz se demostró empíricamente (Bernardo, 2010) que la covariancia entre los desvíos de dominancia era positiva y que la covariancia de los valores de mejora era pequeña. Ambos términos de covariancia implican la existencia de variabilidad genotípica oculta que puede ser expuesta conforme disminuya el desequilibrio gamético o de ligamiento.

El término de covariancia entre los valores genotípicos de pares de loci ($\text{COV}_{(Gi,Gj)}$) por efecto del desequilibrio puede resultar mayor o menor que cero. Dicho de otra manera, mientras exista el desequilibrio gamético la

variancia genotípica de la población puede ser mayor o menor respecto a la variancia genotípica bajo la condición de equilibrio de ligamiento. Cualquiera sea la fase de desequilibrio de ligamiento (en acoplamiento o en repulsión), la variancia de dominancia disminuirá a medida que disminuya el desequilibrio. La variancia aditiva, por el contrario, será menor cuando existe desequilibrio de ligamiento en fase de repulsión y será mayor cuando exista desequilibrio de ligamiento en fase de acoplamiento. Las expresiones que permiten calcular el sesgo de la variancia aditiva y de dominancia en poblaciones en desequilibrio de ligamiento se presentan en la Tabla 5.12 y en la Figura 5.6.

Variancias	Sin ligamiento	Con ligamiento	
		Fase de acoplamiento	Fase de repulsión
Aditiva	σ_A^2	$\sigma_A^2 + [(1-2c)/2]\sigma_A^2$	$\sigma_A^2 - [(1-2c)/2]\sigma_A^2$
Dominante	σ_D^2	$\sigma_D^2 + [(1-2c)/2]^2\sigma_D^2$	

Tabla 5.12.

Efecto del coeficiente de recombinación (c) sobre las variancias aditivas y de dominancia en una población en desequilibrio de ligamiento.

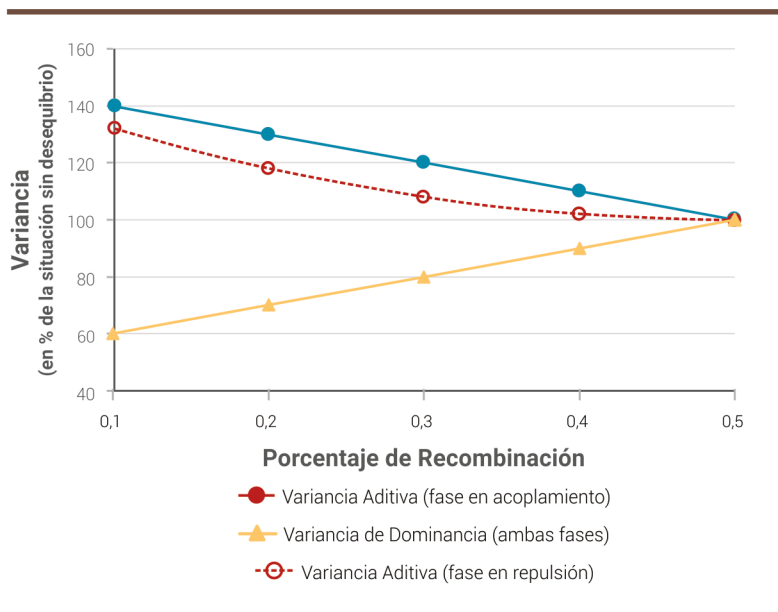


Figura 5.6.

Modificaciones de las variancias aditivas y de dominancia para diferentes coeficientes de recombinación (c) respecto a la condición de equilibrio gamético.

La existencia de desequilibrio de ligamiento tiene implicancias prácticas para la selección. Es frecuente que en poblaciones de cría que se generan por cruzamiento entre poblaciones con diferencias de frecuencias génicas

importantes (por ejemplo entre una población adaptada y otra exótica, o entre una población élite y una población con escaso mejoramiento genético previo), la respuesta a la selección que puede esperarse será menor, precisamente porque en tales tipos de poblaciones el desequilibrio que existirá más probablemente sea en fase de repulsión y menor será la variancia aditiva capitalizable por selección. Por una parte puede resultar conveniente efectuar varias generaciones de apareamiento aleatorio a fin de que el desequilibrio de ligamiento se reduzca y permita “liberar” variabilidad aditiva oculta e incrementar la respuesta esperada a la selección.

Por otra parte cuando el ligamiento es en fase de repulsión, será necesario emplear mayores tamaños muestrales, puesto que los genotipos deseados estarán presentes en menor frecuencia a la esperada. Por el contrario, el ligamiento de genes favorables en fase de acoplamiento constituye una situación ventajosa en términos de requerimientos de muestreo y de respuesta esperada a la selección, porque esos genotipos estarán en una frecuencia mayor a la esperada en condiciones de equilibrio de ligamiento. Es muy común en los programas de mejoramiento la utilización de poblaciones de cría que resultan del cruzamiento entre materiales elite y en consecuencia el desequilibrio esperable será, probablemente, en fase de acoplamiento.

5.9. Heredabilidad

Knight (1948) definió a la heredabilidad como la fracción de la variancia observable que se relaciona con la herencia. El término tiene dos acepciones: en sentido amplio y en sentido estricto.

La heredabilidad en sentido amplio (H) tiene en cuenta toda la variación causada por los distintos tipos de acción génica en relación con la suma de la variabilidad genotípica y la variabilidad causada por el ambiente.

$$H = \sigma_G^2 / [\sigma_G^2 + \sigma_E^2] = (\sigma_A^2 + \sigma_D^2 + \sigma_I^2) / (\sigma_A^2 + \sigma_D^2 + \sigma_I^2 + \sigma_E^2)$$

La heredabilidad en sentido estricto (h^2) solo considera la proporción de la variancia genotípica fijable (aditiva) respecto de la variancia observable o fenotípica. La heredabilidad en sentido estricto expresa en qué proporción las diferencias observables (fenotípicas) entre los individuos de una generación puede transmitirse a la siguiente y es por lo tanto la que ofrece mayor valor práctico en programas de selección.

Recordemos que en organismos diploides y aloploiploides solo la acción génica aditiva es transmisible entre generaciones puesto que durante el proceso de gametogénesis las interacciones intraalélicas (dominancia)

e interalélicas (epistasias) se suprimen por el efecto de la reducción a la condición haploide o se rompen parcialmente por la recombinación genética, respectivamente. Los efectos de dominancia y epistasias solo son parcialmente transmisibles en la medida que los individuos de la siguiente generación recreen las interacciones intraalélicas (porque ofrecen alguna ventaja en la selección de la población original) y que el ligamiento entre loci cercanos que llevan combinaciones alélicas favorables sea suficiente para prevenir la ocurrencia de recombinación genética.

En mejoramiento genético la heredabilidad no solo puede expresarse sobre la base de individuos (animales, plantas), sino también sobre la base de familias.

En el caso de que la unidad de referencia sean plantas individuales de la población de referencia (es decir, que consideramos la variabilidad fenotípica entre plantas sin tener en cuenta las estructuras familiares) la heredabilidad en sentido estricto será:

$$h^2 = \sigma_A^2 / [(\sigma_A^2 + \sigma_D^2 + \sigma_I^2) + \sigma_E^2],$$

donde σ_E^2 es la suma de la variancia ambiental y la de interacción genotipo x ambiente, es decir:

$$h^2 = \sigma_A^2 / [(\sigma_A^2 + \sigma_D^2 + \sigma_I^2) + (\sigma_{GxE}^2 + \sigma^2)]$$

Esta expresión, conceptualmente correcta, en la práctica resulta imposible de estimar con precisión ya que como la unidad de referencia son plantas individuales, no hay posibilidad de evaluarlas en diferentes ambientes (lo que impide estimar σ_{GxE}^2) ni en más de una parcela (lo que impide estimar σ^2).

En el caso de una especie de multiplicación agámica, si efectuáramos una multiplicación clonal de los individuos podrían evaluarse en experimentos con r repeticiones en un único ambiente, conceptualmente la heredabilidad en sentido amplio (H) podría estimarse como:

$$H = \sigma_G^2 / [(\sigma_G^2) + (\sigma_{GxE}^2 + \sigma^2/b)]$$

Pero desde el punto de vista práctico, al evaluar en único ambiente, no es posible estimar σ_{GxE}^2 , término que resultará confundido con la estimación de σ_G^2 . En estas condiciones solo será posible obtener estimaciones sesgadas de σ_G^2 que terminarán sobreestimando la verdadera heredabilidad por cuanto $V_G / (V_G + V_e/r)$, donde V_G estima $\sigma_G^2 + \sigma_{GxE}^2$.

La sobreestimación de la heredabilidad resulta del sesgo en el numerador de la expresión.

Cuando se evalúan ciertos tipos de familias, y si realizáramos el experimento con r repeticiones en a ambientes, la heredabilidad en el sentido estricto sobre la base de medias familiares sería:

$$\sigma_A^2 / [(\sigma_G^2) + (\sigma_{G \times E}^2/a + \sigma^2/ar)]$$

y la estimación:

$$V_A / (V_G + V_{G \times E} / a + V_e/ar), \text{ donde } V_A \text{ estima } \sigma_A^2, V_G \text{ a } \sigma_G^2 \text{ y } V_{G \times E} \text{ a } \sigma_{G \times E}^2$$

Nótese que en este caso V_G y $V_{G \times E}$ son estimadores insesgados de σ_G^2 y de $\sigma_{G \times E}^2$, respectivamente y que a medida que aumenta la magnitud de la evaluación de los genotipos en términos de número de ambientes y de número de repeticiones, se incrementa la estimación de la heredabilidad en sentido estricto.

Si bien ciertos caracteres tienden a poseer heredabilidades que podrían catalogarse como bajas, moderadas o altas, estos parámetros responderán también a las características de determinada población de referencia (frecuencias génicas, y por lo tanto variancias aditivas, de dominancia y de epistasia). Otro factor que incide sobre la heredabilidad, como se presentará más adelante, es si la unidad de referencia son medias de familias o son valores fenotípicos de plantas individuales. Consecuentemente la heredabilidad no es una constante biológica, sino que depende de la población de referencia y de las características experimentales y modalidades de evaluación.

5.9.1. Error estándar de las estimaciones de heredabilidad

Klei y Tsuruta (2008) presentaron un método de estimación del error estándar de un estimador de heredabilidad. Considerando que:

V_A es el estimador de la variancia aditiva,

V_E el estimador de la variancia residual,

V_P el estimador de la variancia fenotípica,

y que $V_P = V_A + V_E$

Entonces el error estándar de la heredabilidad será:

$$ES(h^2) = (V_A/V_P) \{ [V_{VA}/V_A^2] + [V_{VP}/V_P^2] - 2[Cov_{(VA,VP)}/V_A V_P] \}$$

Donde V_{VA} y V_{VP} son las variancias de los estimadores de la variancia aditiva y la variancia fenotípica, respectivamente.

5.10. Utilidad de conocer los diferentes componentes de variancia genotípica en el mejoramiento

Una vez descriptos los componentes de la variancia genotípica y de interacción genotipo x ambiente, es importante determinar la utilidad o la capacidad de aprovecharlos en programas de mejoramiento. Dickerson (1963) resumió sus conclusiones al respecto (Tabla 5.13).

Componente de la variancia genotípica	Entre clones, entre híbridos de líneas homocigotas	Entre líneas homocigotas de especies autóгамas	Entre progenies interpoblacionales	Dentro de una población segregante
σ_A^2	Total	Total	Total	Total
$\sigma_{AA}^2, \sigma_{AAA}^2, \sigma_{AAAA}^2$	Total	Total	Parcial	Parcial
$\sigma_D^2, \sigma_{DD}^2, \sigma_{DDD}^2, \sigma_{ADD}^2 \dots$	Total	Nulo	Parcial	Nulo
σ_{GxE}^2	Nulo	Nulo	Nulo	Nulo

Tabla 5.13.

Grado de aprovechamiento de los componentes de variancia genotípica en la selección sobre diferentes clases de poblaciones o genotipos. Adaptado de Dickerson (1963).

Conocer la variancia fenotípica y sus componentes causales así como de la heredabilidad tiene implicancias prácticas en diferentes aspectos de un programa de mejoramiento. Por una parte, porque existirá una proporcionalidad directa entre la magnitud de la heredabilidad y la respuesta esperada a la selección. Si la variabilidad observada entre individuos o entre medias de familias no respondiera básicamente a la existencia de diferencias heredables entre ellos, solo puede esperarse una respuesta nula a la selección. Por otra parte, la magnitud de la evaluación tiene un efecto muy importante sobre el denominador de la fórmula de heredabilidad ya que en la medida que aumente el número de repeticiones y de ambientes, la variancia fenotípica se reducirá.

Este efecto puede observarse en la Tabla 5.14; en ella se asume que las variancias verdaderas (o paramétricas) fueran 150, 200 y 300 para la genotípica, genotipo x ambiente y ambiental, respectivamente. De la observación de la tabla se concluye que para un mismo nivel de recursos (parcelas) para evaluar un genotipo, resulta más conveniente ampliar el número de ambientes que de repeticiones. Por ejemplo, si se dispone de recursos limitados para evaluar un genotipo en un total de ocho parcelas o unidades experimentales, sería preferible hacerlo en cuatro ambientes con dos repeticiones

y no en dos ambientes de cuatro repeticiones, puesto que la heredabilidad en el primer caso resultaría mayor que en el segundo (0,63 vs. 0,52) y en consecuencia la respuesta a la selección sería mayor.


N.º de Ambientes (a)	N.º de repeticiones (r)	V_G	$V_{G \times E} / a$	V_e / ar	V_P	H
1	1	150	200	300	650	0,23
1	2	150	200	150	500	0,30
1	3	150	200	100	450	0,33
1	4	150	200	75	425	0,35

2	1	150	100	150	400	0,38
2	2	150	100	75	325	0,46
2	3	150	100	50	300	0,50
2	4	150	100	37,5	287,5	0,52

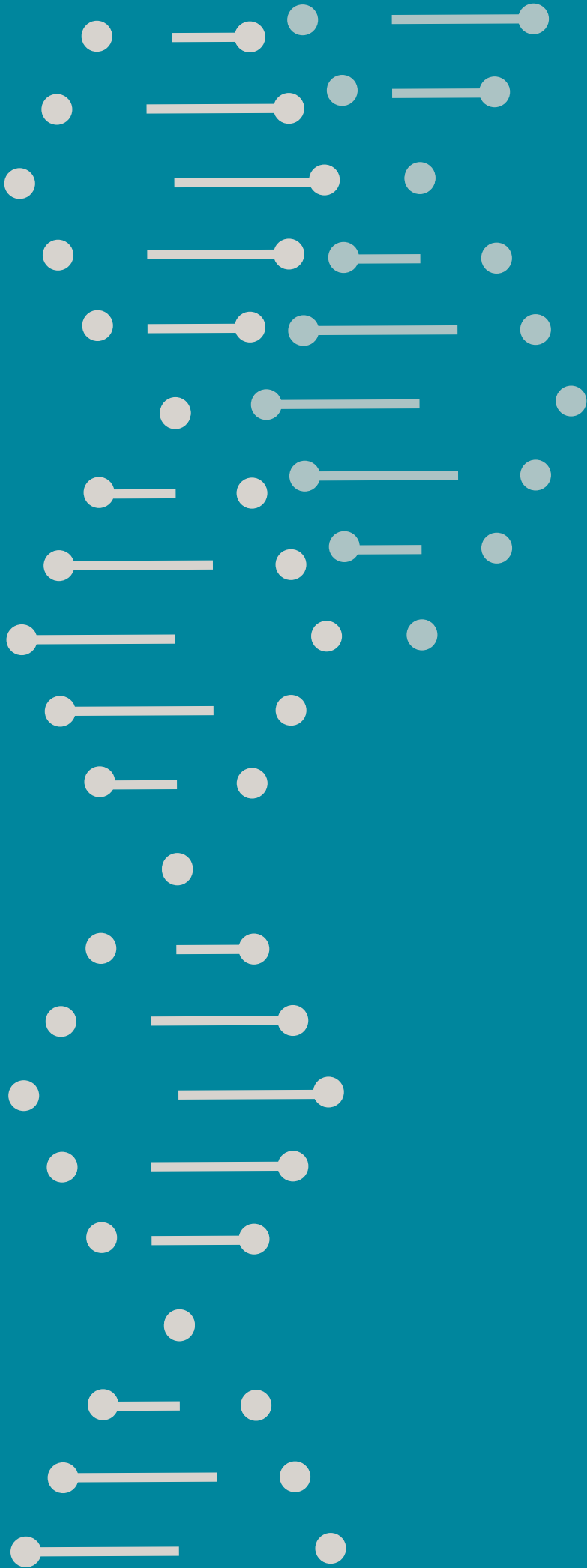
3	1	150	67	100	317	0,47
3	2	150	67	50	267	0,56
3	3	150	67	33	250	0,60
3	4	150	67	25	242	0,62

4	1	150	50	75	275	0,55
4	2	150	50	37,5	237,5	0,63
4	3	150	50	25	225	0,67
4	4	150	50	18,7	218,7	0,69

Tabla 5.14. Efecto de modificaciones en el número de ambientes de evaluación y de repeticiones dentro de ambientes sobre la heredabilidad en sentido amplio.



Cuando la población a mejorar posee poca variabilidad genética aditiva, una manera posible de incrementar la heredabilidad es emplear el tipo de familias que exponga el mayor porcentaje posible de la variancia genética de la población, por ejemplo líneas recombinantes o haploides duplicados. El efecto relativo (Tabla 5.14) de que para un mismo nivel de recursos se modifique su asignación entre número de repeticiones por ambiente y número de ambientes depende también de qué unidad de referencia se toma para definir la heredabilidad en sentido estricto. Debe examinarse caso por caso según la población que se trate.





Medidas de parecido entre parientes

6.1. Causas del parecido entre parientes

Tipos de parientes

Al igual que la media y la variancia fenotípicas, el parecido entre parientes es el resultado de causas genéticas y ambientales. Para que dos o más individuos sean considerados parientes deben poseer al menos un progenitor o ancestro común en su genealogía. El concepto de identidad alélica admite dos situaciones: la de alelos idénticos en estado y la de alelos idénticos por descendencia. En ambos casos nos referimos a alelos que tienen la misma estructura, pero en el segundo la condición de identidad ocurre porque los alelos de dos o más individuos son copias de un mismo alelo ancestral (procedente de generaciones anteriores). En el caso de individuos

emparentados, existe una probabilidad mayor que cero de haber recibido alelos idénticos por descendencia de algún ancestro en común. Expresado en otras palabras, dos o más individuos solo están emparentados si poseen en común copias exactas de al menos parte del total de sus genes, provenientes de la réplica del ADN de uno o más ancestros comunes. La cuantificación de la probabilidad promedio de que los alelos en cada locus de un individuo perteneciente a una población determinada sean producto de la replicación del ADN de algún ancestro es la endogamia promedio, que se expresa mediante el coeficiente de endogamia (F) de Wright (Wright, 1922).

Además de esa causal genética, el ambiente también contribuye al parecido entre individuos emparentados cuando conviven en un mismo ambiente. El de los mellizos es un caso extremo, puesto que estos individuos, además de compartir los dos progenitores, comparten el mismo ambiente intrauterino, la misma habilidad materna, etc.

Tomando en cuenta las genealogías indicadas en la Figura 6.1, podemos reconocer entre los individuos diferentes tipos de parentesco.

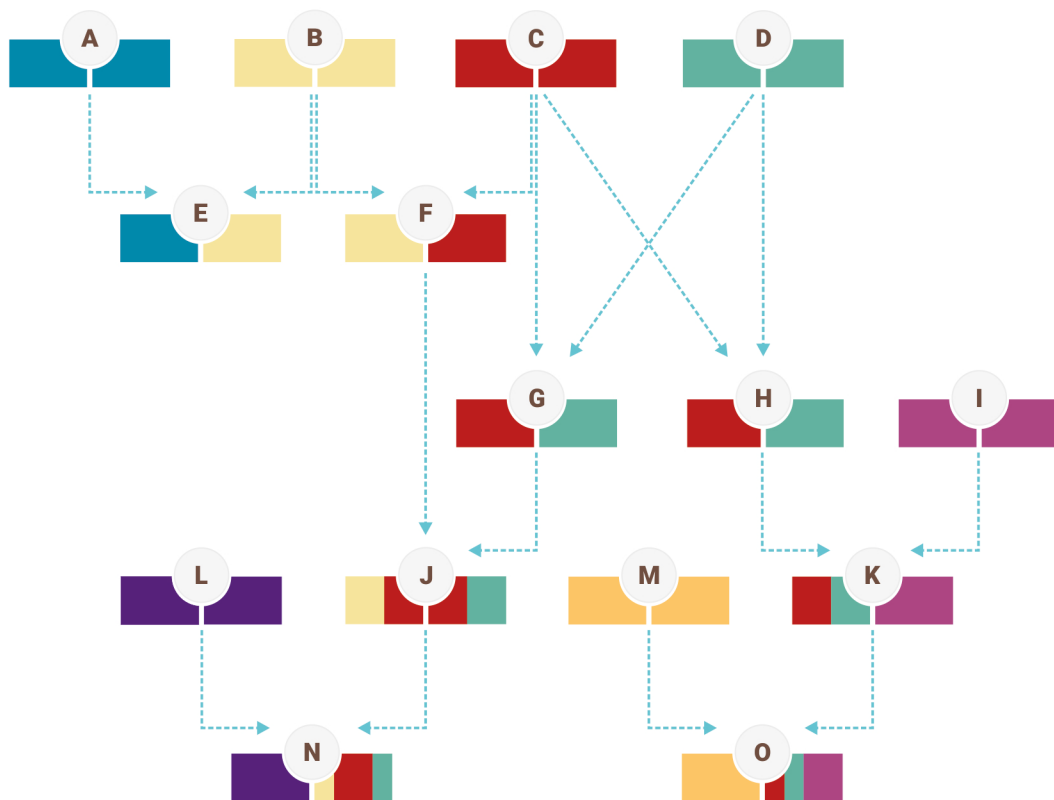


Figura 6.1. Representación de la genealogía hipotética de individuos de una población. Ver texto.

A manera de ejemplo:

- A con E y B con E tienen una relación padre-progenie.
- A y B con E tienen una relación padre promedio-progenie.
- E con F son medios hermanos por tener a B como padre común.
- F con G son medios hermanos por tener a C como padre común.
- G y H son hermanos completos por tener ambos padres, C y D, en común.
- K y J son primos hermanos.
- G con K y H con J guardan una relación de parentesco tío-sobrino.
- N y O son *primos hermanos segundos* por ser hijos, cada uno de ellos, de dos primos hermanos (J y K).

La genealogía indicada en la Figura 6.2 corresponde al parentesco entre primos dobles.

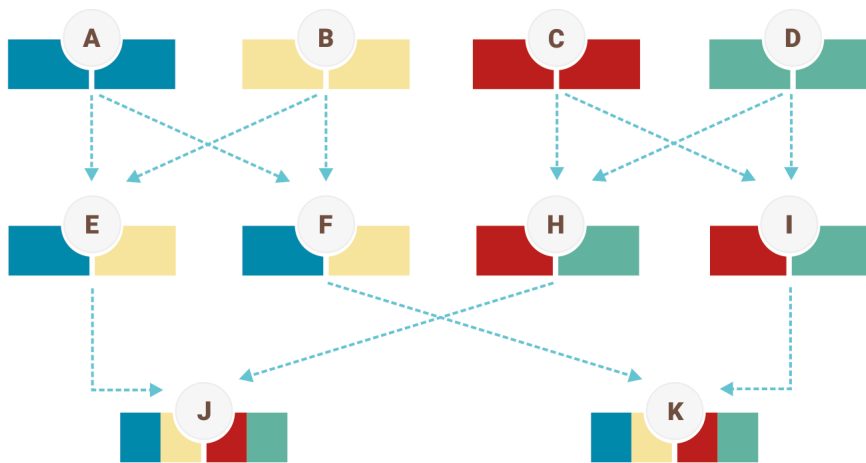


Figura 6.2.
Diagrama que representa la relación de parentesco primos dobles..

Los individuos J y K son primos dobles entre sí, por cuanto uno de sus respectivos progenitores es tío del otro. En este caso E es tío de K e I es tío de J.

6.2. Cuantificación del parecido entre parecidos

Para cuantificar el parecido entre parientes se utilizan medidas de identidad por descendencia, tales como el coeficiente de Malecot (1948) o de coancestría (*coancestry* o *kinship*) siguiendo a Falconer y Mackay (2001), o de coascendencia (*parentage* según Kempthorne, 1957), el coeficiente de

parentesco (*relationship*, Kempthorne, 1957) y el coeficiente de dominancia (Kempthorne, 1957).

6.2.1. Coeficiente de endogamia o de endocría. Identidad alélica por descendencia a nivel de individuos

Es la probabilidad de que los dos alelos en un locus tomado al azar en un individuo diploide Z sean idénticos por descendencia (Wright, 1922). Para el cálculo del coeficiente de endocría (F_Z) es necesario conocer la genealogía desde Z hasta llegar a uno o más ancestros comunes de sus dos progenitores, ya sea porque se dispone de la información completa del pedigrí del individuo Z, o porque este procede de un sistema de apareamiento aplicado en forma recurrente durante varias generaciones, lo que se denomina sistema de endocría. El coeficiente F representa el porcentaje esperado de loci de un individuo donde se espera que sus dos alelos sean idénticos por descendencia (y por lo tanto homocigotas). Visto a nivel poblacional o generacional, F expresa el porcentaje promedio de individuos que se espera sean homocigotas para alelos idénticos por descendencia en un locus tomado al azar.

La aplicación del método de los coeficientes de sendero para el cálculo de F_Z se presenta en la Figura 6.3.

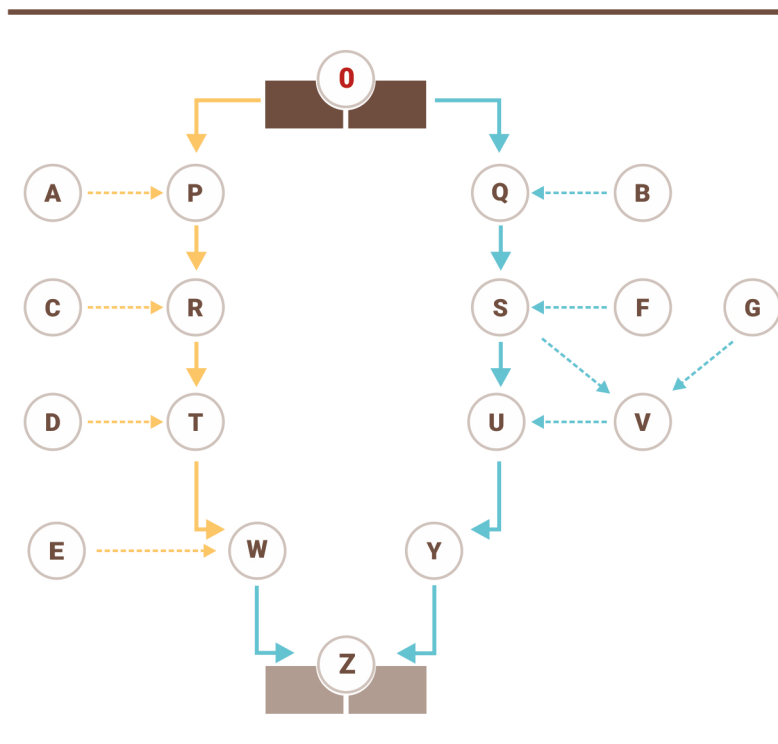


Figura 6.3. Diagrama genealógico correspondiente al individuo Z hasta el antecesor común (O) de sus ancestros (P y Q).

Deben considerarse todos los senderos o caminos posibles que conectan entre sí a ambos padres de Z (W e Y), pasando por todos los ancestros comunes (en el diagrama, O). La probabilidad de que un alelo de un individuo diploide llegue con una copia a su descendencia inmediata es $\frac{1}{2}$. El F_Z se calculará entonces como la probabilidad final de que el alelo ancestral se haya transferido desde el ancestro común (O) hasta Z a través de los diferentes caminos o senderos que conecten a sus padres (W e Y).

Así en el caso del pedigrí que se presenta la Figura 6.3, hay dos caminos posibles que partiendo de uno de los progenitores de Z (W) y pasando por el ancestro común (O) llegan hasta el otro progenitor de Z (Y):

WTRPOQSUY y WTRPOQSVUY.

En estos senderos hay 9 y 10 genotipos involucrados, respectivamente, que conectan Z partiendo de uno de los padres y regresando a través del otro padre hasta Z. Por lo tanto el coeficiente de endocria del individuo Z será:

$$F_Z = (\frac{1}{2})^9 + (\frac{1}{2})^{10} = 1/512 + 1/1024 = 0,00293$$

El coeficiente de endogamia en una población endocriada puede estimarse a partir de la proporción de individuos heterocigotas en la población (H_{IP}) en relación con la proporción esperada en condiciones de apareamiento aleatorio (H_{AA}):

$$(H_{IP} - H_{AA})/H_{AA} = 1 - (H_{IP}/H_{AA}) = F$$

6.2.2. Coeficiente de coancestría. Identidad alélica por descendencia a nivel de pares de individuos

Dados dos individuos, X e Y, su coeficiente de Malecot (1948) o de coancestría (*coancestry*, *kinship*, Falconer y Mackay, 2001) o de coascendencia (Kempthorne, 1957), (f_{xy}), es la probabilidad de que un alelo tomado al azar en un locus de X y otro alelo tomado al azar del mismo locus, pero de Y, resulten ser idénticos por descendencia (es decir que sean copias de un alelo en algún ancestro común). Se expresa simbólicamente como:

$$f_{XY} = P(x \equiv y)$$

A nivel poblacional, el coeficiente de coancestría representa la proporción de genes de un individuo que son idénticos por descendencia a los de otro individuo.

Calcularemos, como ejemplo, el coeficiente de coancestría de los individuos X e Y, con genotipos x_1x_2 e y_1y_2 , respectivamente, de acuerdo a la genealogía que se presenta en el Cuadro 6.1.

Generación	Individuos			
0	P	Q	R	S
1	X (x ₁ ,x ₂)		Y (y ₁ ,y ₂)	
2	Z			

Cuadro 6.1.
Genealogía hipotética para tres generaciones. En la generación base el cruzamiento entre los individuos P y Q genera el hijo X, y el cruzamiento entre R y S genera el hijo Y. Los individuos X e Y de la generación 1 engendran el individuo Z en la generación 2.

El individuo X es hijo de P y Q, en tanto que Y es hijo de R y S. A su vez Z, es hijo de X y de Y. Los subíndices 1 o 2 se refieren a los alelos de los individuos X e Y.

El coeficiente de coancestría entre X e Y será la probabilidad de que un alelo tomado al azar de X y otro de Y resulten idénticos por descendencia. Del cruzamiento entre X e Y pueden producirse en total cuatro combinaciones de gametas posibles (cigóticas) (x₁y₁, x₁y₂, x₂y₁, x₂y₂) y por lo tanto debemos calcular la probabilidad de que en cada una de esas cuatro combinaciones gaméticas los alelos involucrados sean idénticos por descendencia. Es decir, que debemos considerar las probabilidades de que x₁ sea idéntico a y₁, que x₁ sea idéntico a y₂, que x₂ sea idéntico a y₁ o que x₂ sea idéntico a y₂. Cada uno de estos sucesos tiene una probabilidad P de ocurrencia y como se trata de sucesos independientes, la *probabilidad promedio* final de que ocurra al menos uno de esos sucesos es la suma de las probabilidades de cada uno de ellos dividido por el número de sucesos posibles (4).

De tal manera:

$$f_{XY} = 1/4 [P(x_1 \equiv y_1) + P(x_1 \equiv y_2) + P(x_2 \equiv y_1) + P(x_2 \equiv y_2)]$$

Hallauer (comunicación personal) propuso un método sencillo para calcular el coeficiente de coancestría entre individuos. Conceptualmente:

$$f_{XY} = \sum (P_{x_i y_j}) \cdot P(x_i \equiv y_j)$$

donde:

$P_{x_i y_j}$ es la probabilidad de cada combinación de gametas (cigotas) para X y para Y,

$P(x_i \equiv y_j)$ es la probabilidad de identidad por descendencia de cada combinación de gametas (cigotas) entre X e Y.

Por ejemplo, $P(x_1 \equiv y_1)$ es la probabilidad de que un alelo de X (x₁) recibido por vía materna sea idéntico por descendencia a un alelo de Y (y₁), también



recibido por vía materna. $P(x_1 \equiv y_2)$ es la probabilidad de que el alelo (x_1) de X recibido por vía materna sea idéntico por descendencia al alelo (y_2) de Y, recibido por vía paterna y así sucesivamente con las dos alternativas restantes.

Podemos aplicar este razonamiento para calcular el coeficiente de coancestría para cada tipo de parentesco.

6.2.2.1. Coeficiente de coancestría en familias de medios hermanos

En esta relación de parentesco los individuos X e Y tienen en común uno de los progenitores; en el Cuadro 6.2 que se presenta tal progenitor común es Q, cuyo genotipo es q_1q_2 .

Generación	Individuos		
0	P (p_1p_2)	Q (q_1q_2)	S (s_1s_2)
1	X (x_1x_2)		Y (y_1y_2)
2	Z		

Cuadro 6.2. Genealogía hipotética para tres generaciones. En la generación base el cruzamiento entre los individuos P y Q genera el hijo X, y el cruzamiento entre Q y S genera el hijo Y. Los individuos X e Y de la generación 1 resultan medios hermanos por compartir el progenitor Q.

- ¿Cuáles son las combinaciones gaméticas (cigotas) posibles para el individuo X?

$$p_1q_1 \quad p_1q_2 \quad p_2q_1 \quad p_2q_2$$

- ¿Cuáles son las combinaciones gaméticas (cigotas) posibles para el individuo Y?

$$s_1q_1 \quad s_1q_2 \quad s_2q_1 \quad s_2q_2$$

Es decir que cuando se toman al azar un par de individuos medios hermanos, X e Y, tendremos los sucesos (o pares de genotipos) posibles que se presentan en el Cuadro 6.3.

p_1q_1	s_1q_1	p_1q_2	s_1q_1	p_2q_1	s_1q_1	p_2q_2	s_1q_1
p_1q_1	s_1q_2	p_1q_2	s_1q_2	p_2q_1	s_1q_2	p_2q_2	s_1q_2
p_1q_1	s_2q_1	p_1q_2	s_2q_1	p_2q_1	s_2q_1	p_2q_2	s_2q_1
p_1q_1	s_2q_2	p_1q_2	s_2q_2	p_2q_1	s_2q_2	p_2q_2	s_2q_2

Cuadro 6.3. Pares de combinaciones cigóticas posibles entre individuos medios hermanos.

Estas dieciséis combinaciones son todas igualmente probables, pero en solo ocho de ellas (destacadas en color y negrita) se comparten alelos idénticos por descendencia. Para cada combinación de individuos indicados en **negrita** obtendremos la siguiente probabilidad de identidad por descendencia (tomando por ejemplo la de la izquierda de la primera fila en el Cuadro 6.3):

$$(\frac{1}{4})[P(p_1 \equiv s_1) + P(p_1 \equiv q_1) + P(q_1 \equiv s_1) + P(q_1 \equiv q_1)] = (\frac{1}{4})(0 + 0 + 0 + 1) = 1/4$$

Por lo tanto $f_{XY} = (8/16) \cdot (1/4) = 1/8$

Podemos aplicar la misma metodología para determinar el coeficiente de coancestría de hermanos completos.

6.2.2.2. Coancestría entre hermanos completos

En esta relación de parentesco (Cuadro 6.4) los individuos X e Y tienen en común sus dos progenitores (P y Q).

Generación	Individuos			
0	P (p₁p₂)	Q (q₁q₂)	Q (q₁q₂)	P (p₁p₂)
1	X (x ₁ x ₂)		Y (y ₁ y ₂)	
2	Z			

Cuadro 6.4. Genealogía hipotética para tres generaciones. En la generación base el cruzamiento entre los individuos P y Q genera los hijos X e Y. Los individuos X e Y de la generación 1 resultan ser hermanos completos por compartir sus dos progenitores.

- ¿Cuáles son las combinaciones gaméticas (cigotas) posibles para el individuo X?

$$p_1q_1 \quad p_1q_2 \quad p_2q_1 \quad p_2q_2$$

- ¿Cuáles son las combinaciones gaméticas (cigotas) posibles para el individuo Y?

$$p_1q_1 \quad p_1q_2 \quad p_2q_1 \quad p_2q_2$$

Es decir, que cuando se toman al azar un par de individuos que son hermanos completos entre sí tendremos los siguientes sucesos posibles (Cuadro 6.5):

p₁q₁	p₁q₁	p₁q₂	p₁q₁	p₂q₁	p₁q₁	p₂q₂	p₁q₁
p₁q₁	p₁q₂	p₁q₂	p₁q₂	p₂q₁	p₁q₂	p₂q₂	p₁q₂
p₁q₁	p₂q₁	p₁q₂	p₂q₁	p₂q₁	p₂q₁	p₂q₂	p₂q₁
p₁q₁	p₂q₂	p₁q₂	p₂q₂	p₂q₁	p₂q₂	p₂q₂	p₂q₂

Cuadro 6.5. Pares de combinaciones cigóticas posibles entre individuos hermanos completos.

Las 16 combinaciones son todas igualmente probables, pero esta vez 12 de ellas (destacadas en color) comparten alelos idénticos por descendencia, en número de dos alelos (cuatro combinaciones) o un alelo (ocho combinaciones).

Para cada una de las cuatro combinaciones de individuos indicados en color que comparten dos alelos idénticos obtendremos la siguiente probabilidad de identidad por descendencia (tomando como ejemplo p_1q_1 y p_1q_1):

$$(\frac{1}{4})[P(p_1 \equiv p_1) + P(p_1 \equiv q_1) + P(q_1 \equiv p_1) + P(q_1 \equiv q_1)] = (\frac{1}{4})(1 + 0 + 0 + 1) = \frac{1}{2}$$

Para cada una de las ocho combinaciones de individuos indicados en color que comparten solamente un alelo idéntico obtendremos la siguiente probabilidad de identidad por descendencia (tomando como ejemplo p_1q_1 y p_1q_2):

$$(\frac{1}{4})[P(p_1 \equiv p_1) + P(p_1 \equiv q_2) + P(q_1 \equiv p_1) + P(q_1 \equiv q_2)] = \frac{1}{4}(1 + 0 + 0 + 0) = \frac{1}{4}$$

Por lo tanto $f_{XY} = (4/16) \cdot (1/2) + (8/16) \cdot (1/4) = \frac{1}{4}$

6.2.2.3. Coancestría entre progenie-progenitor

En esta relación de parentesco (Cuadro 6.6) el individuo X es hijo de P.

Generación	Individuos	
0	P (p_1p_2)	Q (q_1q_2)
1	X (x_1x_2)	

Cuadro 6.6. Genealogía hipotética para dos generaciones. En la generación base el cruzamiento entre los individuos P y Q genera el hijo X.

- ¿Cuáles son las combinaciones gaméticas (cigotas) posibles para el individuo X?

$$p_1q_1 \quad p_1q_2 \quad p_2q_1 \quad p_2q_2$$

Es decir, que cuando se toman al azar un individuo X y el progenitor P las posibles combinaciones de cigotas serán las que se indican en el Cuadro 6.7.

p_1p_2	p_1q_1
p_1p_2	p_1q_2
p_1p_2	p_2q_1
p_1p_2	p_2q_2

Cuadro 6.7. Pares de combinaciones cigóticas posibles entre un individuo y uno de sus progenitores.

Las cuatro combinaciones son entre sí igualmente probables y en todas ellas se comparte un alelo idéntico por descendencia. Para cada una de las cuatro combinaciones de individuos indicados en color que comparten un alelo idéntico obtendremos la siguiente probabilidad de identidad por descendencia (como ejemplo la de la primera fila):

$$\frac{1}{4} [P(p_1 \equiv p_1) + P(p_1 \equiv q_1) + P(p_2 \equiv p_1) + P(p_2 \equiv q_1)] = \frac{1}{4} (1 + 0 + 0 + 0) = \frac{1}{4}$$

Por lo tanto $f_{XY} = (4/4) \cdot (1/4) = 1/4$

6.2.2.4. Coancestría entre tío y sobrino

Los individuos S e Y son hermanos completos y X, la descendencia de S y R, es sobrino del individuo Y (Cuadro 6.8).

Generación	Individuos				
0		P (p ₁ p ₂)	Q (q ₁ q ₂)	P (p ₁ p ₂)	Q (q ₁ q ₂)
1	R (r ₁ r ₂)	S (z ₁ z ₂)		Y (y ₁ y ₂)	
2	X (x ₁ x ₂)				

Cuadro 6.8. Genealogía hipotética para tres generaciones. En la generación base el cruzamiento entre los individuos P y Q genera los hijos S e Y, y el cruzamiento de los individuos R y S genera el individuo X. Los individuos X e Y guardan un parentesco del tipo tío-sobrino.

En términos de los alelos de P y de Q, las combinaciones gaméticas (cigotas) posibles del individuo S serán:

$$p_1q_1 \quad p_1q_2 \quad p_2q_1 \quad p_2q_2$$

Por lo tanto las combinaciones gaméticas (cigotas) posibles para el individuo X serán las que se indican en el Cuadro 6.9.

		1/4	1/4	1/4	1/4
		p ₁	q ₁	p ₂	q ₂
1/2	r ₁	r ₁ p ₁	r ₁ q ₁	r ₁ p ₂	r ₁ q ₂
1/2	r ₂	r ₂ p ₁	r ₂ q ₁	r ₂ p ₂	r ₂ q ₂

Cuadro 6.9. Combinaciones cigóticas posibles para el individuo X, resultante del cruzamiento entre los individuos R y S (ver pedigrí en Cuadro 6.8).

La probabilidad de cada combinación gamética (cigota) de X es 1/8 y esas combinaciones gaméticas (cigotas) en el individuo X serán las indicadas en el Cuadro 6.10.



r_1p_1	r_1q_1	r_1p_2	r_1q_2
r_2p_1	r_2q_1	r_2p_2	r_2q_2

Cuadro 6.10.
Combinaciones cigóticas posibles para el individuo Y, resultante del cruzamiento entre los individuos P y Q (ver pedigrí en Cuadro 6.8).

– ¿Cuáles son las combinaciones gaméticas (cigotas) posibles para el individuo Y?:

$$p_1q_1 \quad p_1q_2 \quad p_2q_1 \quad p_2q_2$$

Es decir, que cuando se toman al azar un individuo X y su tío Y, las posibles combinaciones de cigotas que podrían encontrarse son las del Cuadro 6.11.

p_1q_1	r_1p_1	p_1q_2	r_1p_1	p_2q_1	r_1p_1	p_2q_2	r_1p_1
p_1q_1	r_1q_1	p_1q_2	r_1q_1	p_2q_1	r_1q_1	p_2q_2	r_1q_1
p_1q_1	r_1p_2	p_1q_2	r_1p_2	p_2q_1	r_1p_2	p_2q_2	r_1p_2
p_1q_1	r_1q_2	p_1q_2	r_1q_2	p_2q_1	r_1q_2	p_2q_2	r_1q_2
p_1q_1	r_2p_1	p_1q_2	r_2p_1	p_2q_1	r_2p_1	p_2q_2	r_2p_1
p_1q_1	r_2q_1	p_1q_2	r_2q_1	p_2q_1	r_2q_1	p_2q_2	r_2q_1
p_1q_1	r_2p_2	p_1q_2	r_2p_2	p_2q_1	r_2p_2	p_2q_2	r_2p_2
p_1q_1	r_2q_2	p_1q_2	r_2q_2	p_2q_1	r_2q_2	p_2q_2	r_2q_2

Cuadro 6.11.
Pares de combinaciones cigóticas posibles entre individuos que guardan una relación de parentesco tío-sobrino (ver pedigrí en Cuadro 6.8).

Las 32 combinaciones son todas igualmente probables y la mitad de ellas comparten un alelo idéntico por descendencia.

Para cada una de las 16 combinaciones de individuos indicados en color que comparten dos alelos idénticos obtendremos la siguiente probabilidad de identidad por descendencia (como ejemplo p_1q_1 y r_1p_1):

$$(\frac{1}{4})[P(p_1 \equiv r_1) + P(p_1 \equiv p_1) + P(q_1 \equiv r_1) + P(q_1 \equiv p_1)] = \frac{1}{4} (0 + 1 + 0 + 0) = \frac{1}{4}$$

$$\text{Por lo tanto } f_{XY} = (16/32) \cdot (1/4) = 1/8$$

6.2.2.5. Reglas para calcular el coeficiente de coancestría

Alternativamente, todos estos ejemplos de cálculo de coeficientes de coancestría pueden realizarse aplicando directamente las siguientes reglas

(Falconer y Mackay, 2001) que se presentan sobre la base del esquema general que presenta el Cuadro 6.1.

- La coancestría entre dos individuos en una generación es el promedio de la coancestría entre sus padres.

Aplicado al esquema de más arriba:

$$f_{XY} = (1/4)(f_{PR} + f_{PS} + f_{QR} + f_{QS}) \quad [6.1]$$

- La coancestría entre dos individuos es el promedio de la coancestría entre uno de los dos individuos con cada uno de los padres del otro.

En este caso:

$$f_{XY} = (1/2)(f_{XR} + f_{XS}) \quad [6.2]$$

o bien

$$f_{XY} = (1/2)(f_{YP} + f_{YQ})$$

- La coancestría entre individuos que pertenecen a diferentes generaciones es el promedio de la coancestría de un individuo con los progenitores del otro.

Por ejemplo entre X y R:

$$f_{XR} = (1/2)(f_{PR} + f_{QR})$$

- La coancestría de un individuo consigo mismo es

$$f_{XX} = (1/4)[P(x_1 \equiv x_1) + P(x_1 \equiv x_2) + P(x_2 \equiv x_1) + P(x_2 \equiv x_2)]$$

$$f_{XX} = (1/4)(1 + F_x + F_x + 1)$$

$$f_{XX} = (1/2)(1 + F_x) \quad [6.3]$$

Si no existiera endocria previa, de modo que $F_x = 0$, entonces

$$f_{XX} = (1/4)(1 + 0 + 0 + 1)$$

$$f_{XX} = (1/4)(2) = 1/2$$

- La coancestría entre dos individuos es el coeficiente de endogamia (F) de su progenie:

$$f_{XY} = F_z \quad [6.4]$$

Podemos aplicar estas reglas para calcular el coeficiente de coancestría para los tipos de parientes presentados más arriba.

6.2.2.5.1. Coancestría de familias de medios hermanos

Como vimos más arriba, en esta relación de parentesco los individuos X e Y tienen en común uno de los progenitores. En el Cuadro 6.2 Q, cuyo genotipo es q_1q_2 .

Por la regla [6.1]:

$$f_{XY} = (1/4)(f_{PQ} + f_{PS} + f_{QQ} + f_{QS})$$

Siendo que P, Q y S no están emparentados, aplicando la regla [6.3] y asumiendo que Q no es endogámico:

$$f_{XY} = (1/4)[0 + 0 + (1/2)(1 + F_Q) + 0]$$
$$f_{XY} = (1/4)[0 + 0 + (1/2)(1 + 0) + 0] = 1/8$$

Si asumimos que P, Q y S no están emparentados y que el progenitor común (Q) es endogámico ($F_Q = 1$), entonces:

$$f_{XY} = (1/4)[0 + 0 + (1/2)(1 + 1) + 0] = 1/4$$

6.2.2.5.2. Coancestría de familias de hermanos completos

En esta relación de parentesco los individuos X e Y tienen en común a los dos progenitores. En el Cuadro 6.4 P y Q.

En este caso, aplicando [6.1]:

$$f_{XY} = (1/4)(f_{PP} + f_{PQ} + f_{QP} + f_{QQ})$$

y aplicando [6.3]:

$$f_{XY} = (1/4) [(1/2)(1 + F_P) + 2f_{PQ} + (1/2)(1 + F_Q)]$$

Si asumimos que P y Q no están emparentados ni tienen endocría previa:

$$f_{XY} = (1/4) [(1/2)(1 + 0) + 0 + (1/2)(1 + 0)] = 1/4$$

Si asumimos que P y Q no están emparentados y que ambos progenitores comunes (P, Q) son endogámicos ($F_Q = F_P = 1$), entonces:

$$f_{XY} = (1/4) [(1/2)(1 + 1) + 0 + (1/2)(1 + 1)] = 1/2$$

Si asumimos que P y Q están emparentados y son endogámicos:

$$f_{XY} = (1/2)(1 + f_{PQ})$$

6.2.2.5.3. Coancestría entre tío-sobrino

Corresponde al esquema genealógico del Cuadro 6.8

$$f_{XY} = (1/4)(f_{RP} + f_{RQ} + f_{SP} + f_{SQ})$$

$$f_{XY} = (1/4)[0 + 0 + (1/2)(f_{PP} + f_{PQ})] = (1/4)[(1/2)(1 + F_P) + 0]$$

Si P es no endogámico,

$$f_{XY} = (1/4) [(1/2)(1 + 0) + 0] = 1/8$$

6.2.2.5.4. Coancestría entre progenie y progenitor

Calcularemos la relación de parentesco entre el X y uno de sus progenitores, Q (Cuadro 6.6)

En este caso, aplicando [6.2]:

$$f_{XQ} = (1/2)[f_{XX} + f_{PQ}]$$

y según [6.3]:

$$f_{XQ} = (1/2) [(1/2)(1 + F_X) + f_{PQ}]$$

Si $F_X = 1$, entonces $f_{XQ} = (1/2)[1 + f_{PQ}]$

Si $F_X = 1$ y si P y Q no están emparentados: $f_{XQ} = 1/2$

Si $F_X = 0$ y si P y Q no están emparentados: $f_{XQ} = 1/4$

6.2.2.5.5. Coancestría bajo sistemas regulares de endocría

Bajo estos sistemas los individuos que comparten una misma generación poseen el mismo nivel de endocría, resultante de la repetición del mismo esquema de apareamiento generación tras generación. Desde un punto de vista práctico resulta útil deducir ecuaciones recurrentes para calcular los coeficientes de endogamia y de coancestría y relacionarlos con los respectivos coeficientes en las generaciones previas (Falconer y Mackay, 2001).

Bajo diferentes sistemas de endocría, las ecuaciones recurrentes se presentan en la Tabla 6.1.

Sistema de endocría			
Autofecundación	Apareamiento entre hermanos completos	Apareamiento entre medios hermanos	Retrocruza
$F_t = 1/2 (1 + F_{t-1})$	$F_t = 1/4(1 + 2F_{t-1} + F_{t-2})$	$F_t = 1/8(1 + 6F_{t-1} + F_{t-2})$	$F_t = 1/4 (1 + 2F_{t-1})$

Tabla 6.1.

Ecuaciones para el cálculo del coeficiente de endogamia luego de t generaciones aplicando diferentes sistemas: autofecundación, retrocruzamiento y apareamiento entre medios hermanos y hermanos completos.

6.2.3. Coeficiente de Kempthorne

Este coeficiente fue definido Kempthorne (1954) como la probabilidad promedio de que dos individuos emparentados, X e Y, reciban alelos idénticos por descendencia desde algún ancestro común a partir de uno y de otro progenitor, simbolizadas como ϕ y ϕ' , respectivamente.

6.2.3.1. Caso de hermanos completos

Calcularemos, a modo de ejemplo, el coeficiente de Kempthorne para el caso de individuos hermanos completos, X (genotipo x_1x_2) e Y (genotipo y_1y_2), hijos de los individuos P y Q, de acuerdo a la genealogía que se presenta en el Cuadro 6.4.

De acuerdo al esquema, X e Y son hijos de P y de Q y asumiremos que P y Q no están emparentados ni son endogámicos.

$$r_{xy} = (1/2) (\phi + \phi'),$$

donde:

r_{xy} = es el coeficiente de Kempthorne

$$\phi = P(x_1 \equiv y_1) = P(x_1 \equiv p_1 \equiv y_1) + P(x_1 \equiv p_2 \equiv y_1) = 1/4 + 1/4 = 1/2$$

$$y \ \phi' = P(x_2 \equiv y_2) = P(x_2 \equiv q_1 \equiv y_1) + P(x_2 \equiv q_2 \equiv y_2) = 1/4 + 1/4 = 1/2$$

Por lo tanto:

$$r_{xy} = (1/2)(\phi + \phi') = (1/2)(1/2 + 1/2) = 1/2$$

Obsérvese que $r_{xy} = 2f_{xy}$

6.2.3.2. Caso de medios hermanos

Consideremos el coeficiente de parentesco entre dos individuos, X e Y, que son medios hermanos por tener en común el progenitor Q de acuerdo al Cuadro 6.2.

Según el esquema, X es hijo de P y Q, e Y es hijo de Q y S.

$$r_{xy} = (1/2)(\phi + \phi'),$$

donde:

$$\phi = P(x_1 \equiv y_1) + P(x_2 \equiv y_2) = P(x_1 \equiv q_1 \equiv y_1) + P(x_2 \equiv q_2 \equiv y_2) = 1/2$$

$$y \ \phi' = 0$$

Por lo tanto:

$$r_{xy} = (1/2)(\phi + \phi') = (1/2)(1/2 + 0) = 1/4$$

Obsérvese que $r_{xy} = 2f_{xy}$

Al igual que en el caso de hermanos completos, consideramos que los padres de X e Y no están emparentados ni son endogámicos.

6.2.3.3. Caso de progenie-progenitor

Asumiremos que P y Q no están emparentados, no son endogámicos y que el individuo X es progenie del cruzamiento P x Q, siguiendo el Cuadro 6.6.

De acuerdo al esquema, X es hijo de P y Q. El coeficiente de parentesco entre P y Q será

$$r_{xy} = (1/2)(\phi + \phi'),$$

donde:

$$\phi = P(x_1 \equiv y_1) + P(x_2 \equiv y_2) = 1/2 + 1/2 = 1$$

$$\text{y } \phi' = 0$$

Por lo tanto:

$$r_{xy} = (1/2)(\phi + \phi') = (1/2)(1 + 0) = 1/2$$

6.2.3.4. Caso de parentesco tío-sobrino

Generación	Individuos			
0	P (p ₁ p ₂)	Q (q ₁ q ₂)	P (p ₁ p ₂)	Q (q ₁ q ₂)
1	R (r ₁ r ₂)		S (s ₁ s ₂)	E (e ₁ e ₂)
2	F (f ₁ f ₂)			

Cuadro 6.12.

Genealogía hipotética para tres generaciones. En la generación base el cruzamiento entre los individuos P y Q genera los hijos R y S, y el cruzamiento de los individuos E y S genera el individuo F. Los individuos F y R guardan un parentesco del tipo tío-sobrino.

De acuerdo al Cuadro 6.12 R y S son hermanos completos e hijos de P y Q. F es, a su vez, hijo de S y E, por lo que el parentesco entre R y F es de tío-sobrino

$$r_{xy} = (1/2)(\phi + \phi'),$$

donde:

$$\begin{aligned}\phi &= P(r_1 \equiv f_1) + P(r_2 \equiv f_1) = [P(r_1 \equiv p_1 \equiv s_1 \equiv f_1) + \\ &+ P(r_1 \equiv q_1 \equiv s_1 \equiv f_1)] + [P(r_2 \equiv p_2 \equiv s_2 \equiv f_2) \\ &+ P(r_2 \equiv q_2 \equiv s_2 \equiv f_2)] = \\ &= (1/8 + 1/8) + (1/8 + 1/8) = 1/2 \\ \text{y } \phi' &= 0\end{aligned}$$

Por lo tanto:

$$r_{xy} = (1/2)(\phi + \phi') = (1/2)(1/2 + 0) = 1/4$$

6.3. Coeficiente de dominancia

Este coeficiente (Δ_{xy}) fue definido por Kempthorne (1957) y es la probabilidad de que dos individuos, X e Y, posean el mismo genotipo idéntico por descendencia. A diferencia de los anteriores coeficientes, el de dominancia solo podrá ser distinto de cero si y solo si los individuos X e Y tienen ancestros en común a través de sus dos progenitores.

Consideremos las relaciones que presenta el Cuadro 6.1

$$\text{Aplicando la definición: } \Delta_{XY} = f_{PR} \cdot f_{QS} + f_{PS} \cdot f_{QR} = \phi\phi'$$

Calcularemos, a modo de ejemplo, el coeficiente de dominancia entre individuos hermanos completos X (genotipo x_1x_2) e Y (genotipo y_1y_2) de acuerdo a la genealogía que se presenta en el Cuadro 6.4:

En este caso X e Y son hijos del cruzamiento entre los progenitores P y Q.

Los individuos X e Y pueden tener cualquiera de los siguientes genotipos:

$$p_1q_1$$

$$p_1q_2$$

$$p_2q_1$$

$$p_2q_2$$

Por lo tanto debemos calcular las siguientes probabilidades y sumarlas ya que cualquiera de los cuatro casos cumple la condición exigida en la definición del coeficiente de dominancia:

$$P(x_1 \equiv y_1 \equiv p_1) \times P(x_2 \equiv y_2 \equiv q_1) = (1/2 \cdot 1/2) \times (1/2 \cdot 1/2) = 1/16$$

$$P(x_1 \equiv y_1 \equiv p_1) \times P(x_2 \equiv y_2 \equiv q_2) = (1/2 \cdot 1/2) \times (1/2 \cdot 1/2) = 1/16$$

$$P(x_1 \equiv y_1 \equiv p_2) \times P(x_2 \equiv y_2 \equiv q_1) = (1/2 \cdot 1/2) \times (1/2 \cdot 1/2) = 1/16$$

$$P(x_1 \equiv y_1 \equiv p_2) \times P(x_2 \equiv y_2 \equiv q_2) = (1/2 \cdot 1/2) \times (1/2 \cdot 1/2) = 1/16$$

Es decir que:

$$\Delta_{xy} = 4(1/16) = 1/4$$

Para el caso de hermanos completos recordemos que ϕ y ϕ' eran $1/2$.

Por lo tanto $\Delta_{xy} = (1/2) (1/2) = 1/4$

Entre los coeficientes de parentesco y de dominancia existe la siguiente relación:

$$\Delta_{xy} = \phi \phi' y r_{xy} = (1/2)(\phi + \phi')$$

Por lo tanto $\Delta_{xy} = \phi'(2r_{xy} - \phi')$

En la Tabla 6.2 se presentan los coeficientes de coancestría, de Kempthorne y de dominancia para diferentes parentescos.

Parentesco	f_{xy}	r_{xy}	Δ_{xy}
Padre – Progenie	1/4	1/2	0
Medios Hermanos	1/8	1/4	0
Hermanos Completos	1/4	1/2	1/4
Tío-Sobrino	1/8	1/4	0

Tabla 6.2.
Coeficientes de coancestría (f_{xy}), de Kempthorne (r_{xy}) y de dominancia (Δ_{xy}) para diferentes tipos de parentescos.

6.4. Cálculo de coeficientes de parentesco en esquemas de pedigrí complejos mediante el método tabular

Siguiendo el esquema del pedigrí de interés debe armarse una tabla de doble entrada. La primera fila y primera columna de la tabla se construyen colocando de izquierda a derecha y de arriba hacia abajo a los individuos más antiguos en el pedigrí y siguiendo por los más recientes. Se elabora entonces una tabla o matriz cuadrada y simétrica, de tantas filas o columnas como individuos hay en el diagrama de pedigrís (Freeman, 1988). Los elementos de la diagonal de esa matriz (a_{ii}) son 1 más el coeficiente de

endogamia (F) del individuo que corresponde a la i-ésima fila y columna. Es decir que:

$$1 + F_i = a_{ii}$$

Los elementos de la matriz ubicados por fuera de la diagonal (a_{ij}) son los coeficientes de parentesco de Wright (R_{ij}) de la combinación de individuos correspondientes a la combinación fila/columna. Cuando se desconocen los progenitores de la combinación del i-ésimo y j-ésimo genotipo, el elemento a_{ij} correspondiente será 0; y en el caso de la combinación del i-ésimo genotipo consigo mismo será 1.

A su vez, $R_{ij} = a_{ij}/(a_{ii} a_{jj})^{1/2}$ [6.5]

Los elementos de la tabla se van determinando secuencialmente de la siguiente manera, comenzando con los elementos a_{ii} :

$$a_{ii} = 1 + (1/2)a_{i'j'} \quad [6.6]$$

donde i' y j' son los padres de i , o bien su equivalente:

$$a_{ii} = 1 + (1/2) (\text{elemento de la interacción de los padres de } i) \quad [6.7]$$

Los términos a_{ij} se computan como:

$$a_{ij} = (1/2)(a_{ij'} + a_{ij''})$$

donde j' y j'' son los padres de j , o bien

$$a_{ij} = (1/2) (\text{la suma de los elementos en la } i\text{-ésima fila para los dos padres de } j')$$

Los elementos a_{ji} son idénticos a los elementos a_{ij} , por lo que no son necesarios nuevos cálculos.

Apliquemos estas reglas al esquema de pedigrí que se presenta en la Figura 6.4 (tomado de Freeman):

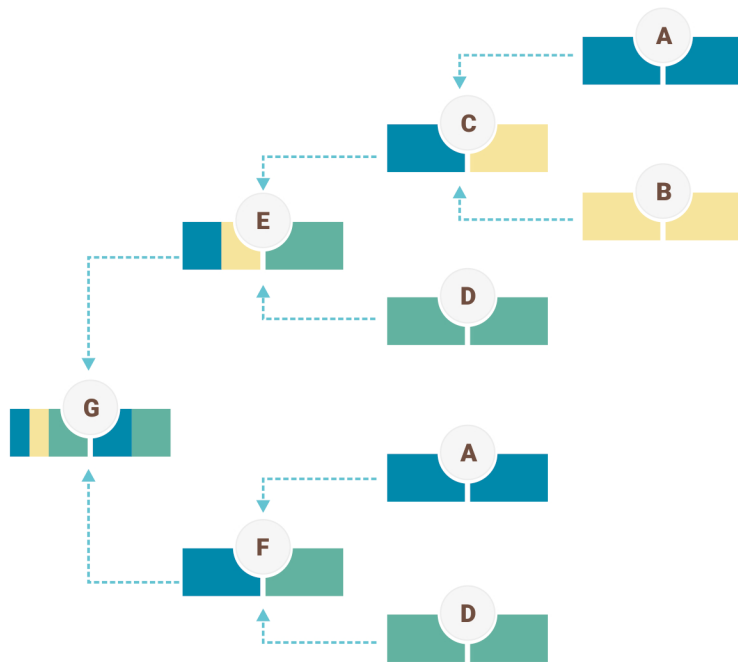


Figura 6.4. Diagrama hipotético de genealogía compleja para ejemplificar la aplicación del método tabular en el cálculo del coeficiente de parentesco

Los elementos $a_{AB'}$, $a_{AD'}$, $a_{BD'}$, a_{CD} y a_{FB} (y sus recíprocos) son iguales a 0 por cuanto no hay información en el pedigrí que permita determinar que esos pares de individuos estén relacionados.

Los elementos $a_{AA'}$, a_{BB} y a_{DD} son iguales a 1, por cuanto no conocemos si tienen endogamia previa (se asume que $F_{A'}$, F_B y F_D son iguales a cero).

Los elementos a_{CC} y a_{FF} pueden computarse aplicando [6.6] o [6.7]:

$$a_{CC} = 1 + (1/2)a_{AB} = 1 + (1/2)(0) = 1 ; F_C = 0$$

$$a_{FF} = 1 + (1/2)a_{AD} = 1 + (1/2)(0) = 1 ; F_F = 0$$

Los restantes elementos:

$$a_{AC} = (1/2)(a_{AA} + a_{AB}) = (1/2)(1+0) = 1/2$$

$$a_{AE} = (1/2)(a_{AC} + a_{AD}) = (1/2)(1/2 + 0) = 1/4$$

$$a_{BC} = (1/2)(a_{BA} + a_{BB}) = (1/2)(0 + 1) = 1/2$$

$$a_{BE} = (1/2)(a_{BC} + a_{BD}) = (1/2)(1/2 + 0) = 1/4$$

$$a_{CE} = (1/2)(a_{CC} + a_{CD}) = (1/2)(1 + 0) = 1/2$$

$$a_{DE} = (1/2)(a_{DC} + a_{DD}) = (1/2)(0 + 1) = 1/2$$

$$a_{DF} = (1/2)(a_{DA} + a_{DD}) = (1/2)(0 + 1) = 1/2$$

$$\begin{aligned}
a_{EF} &= (1/2)(a_{EA} + a_{ED}) = (1/2)(1/4 + 1/2) = 3/8 \\
a_{AF} &= (1/2)(a_{AA} + a_{AD}) = (1/2)(1 + 0) = 1/2 \\
a_{AG} &= (1/2)(a_{AE} + a_{AF}) = (1/2)(1/4 + 1/2) = 3/8 \\
a_{BG} &= (1/2)(a_{BE} + a_{BF}) = (1/2)(1/4 + 0) = 1/8 \\
a_{CG} &= (1/2)(a_{CE} + a_{CF}) = (1/2)(1/2 + 1/4) = 3/8 \\
a_{DG} &= (1/2)(a_{DE} + a_{DF}) = (1/2)(1/2 + 1/2) = 1/2 \\
a_{EG} &= (1/2)(a_{EE} + a_{EF}) = (1/2)(1 + 3/8) = 11/16 \\
a_{FG} &= (1/2)(a_{FE} + a_{FF}) = (1/2)(3/8 + 1) = 11/16 \\
a_{GG} &= 1 + (1/2)(a_{EF}) = 1 + (1/2)(3/8) = 19/16; F_G = 3/16
\end{aligned}$$

Obsérvese que el cómputo lleva un orden según la cronología del diagrama de pedigrí (Tabla 6.3). Los últimos elementos corresponden a los genotipos más recientes y se computan en base a los elementos de los genotipos ancestrales.

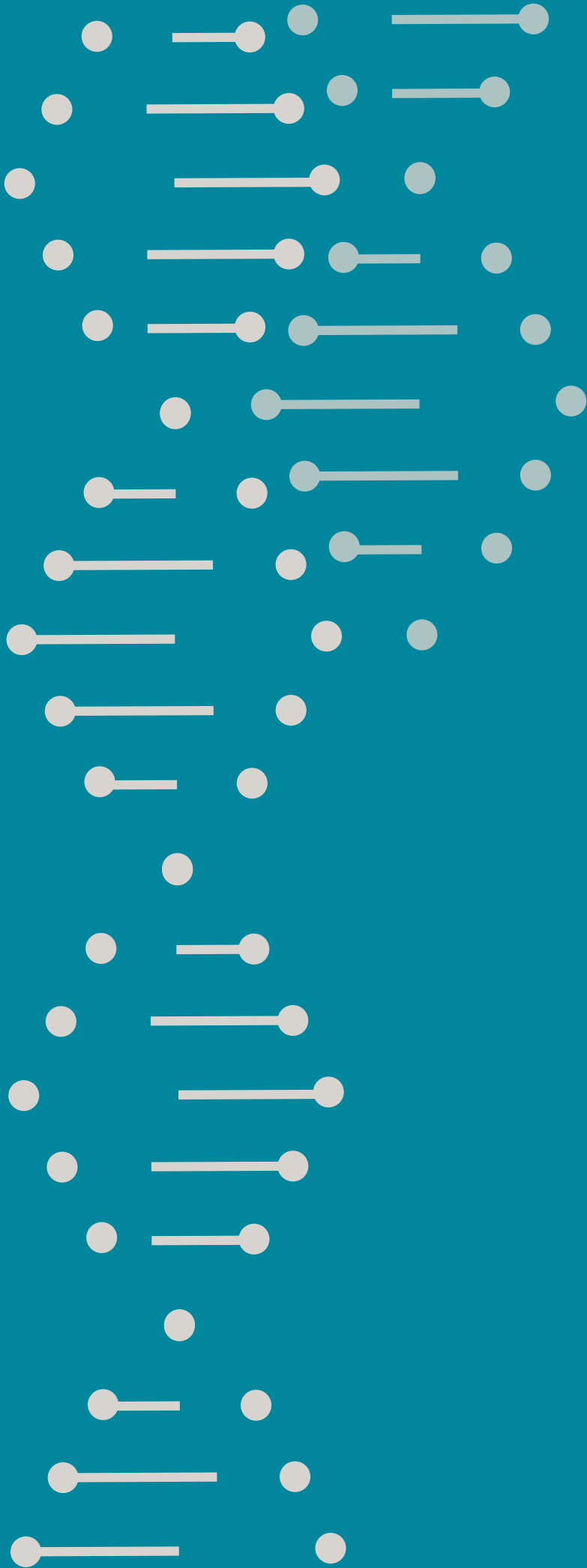
		A	B	C	D	E	F	G
A	1.0	0	1/2	0	1/4	1/2	3/8	
B	0	1.0	1/2	0	1/4	0	1/8	
C	1/2	1/2	1.0	0	1/2	1/4	3/8	
D	0	0	0	1.0	1/2	1/2	1/2	
E	1/4	1/4	1/2	1/2	1.0	3/8	11/16	
F	1/2	0	1/4	1/2	3/8	1.0	11/16	
G	3/8	1/8	3/8	1/2	11/16	11/16	19/16	

Tabla 6.3.

Elementos de la matriz para el cálculo del coeficiente de parentesco aplicando el método tabular al esquema genealógico presentado en la Figura 6.4.

El coeficiente de parentesco de Wright, por ejemplo entre los genotipos F y C, puede calcularse utilizando la expresión [6.5] y utilizando los elementos de la tabla:

$$R_{FC} = a_{FC} / (a_{FF} a_{CC})^{1/2} = (1/4) / [(1)(1)]^{1/2} = 1/4$$



Covariancia entre parientes

7.1. Covariancia genotípica entre individuos y sus componentes

Consideremos los valores genotípicos de dos individuos, X e Y , tomados al azar de una población de referencia, expresados en función de sus valores de mejora (sumatoria de los efectos medios de sus alelos) y los desvíos por dominancia:

$$X_{ij} = \mu + {}^x\alpha_i + {}^x\alpha_j + {}^x\delta_{ij}$$

$$Y_{ij} = \mu + {}^y\alpha_i + {}^y\alpha_j + {}^y\delta_{ij}$$

Por definición, la covariancia genotípica entre X e Y será igual a la esperanza matemática de los productos cruzados:

$$\begin{aligned} \text{Cov}_{G(X,Y)} &= \mathcal{E} ({}^x\alpha_i + {}^x\alpha_j + {}^x\delta_{ij}) ({}^y\alpha_i + {}^y\alpha_j + {}^y\delta_{ij}) = \mathcal{E}({}^x\alpha_i {}^y\alpha_i) + \\ &+ \mathcal{E}({}^x\alpha_i {}^y\alpha_j) + \mathcal{E}({}^x\alpha_i {}^y\delta_{ij}) + \mathcal{E} ({}^x\alpha_j {}^y\alpha_i) + \mathcal{E}({}^x\alpha_j {}^y\alpha_j) + \\ &+ \mathcal{E} ({}^x\alpha_j {}^y\delta_{ij}) + \mathcal{E} ({}^x\delta_{ij} {}^y\alpha_i) + \mathcal{E}({}^x\delta_{ij} {}^y\alpha_j) + \mathcal{E}({}^x\delta_{ij} {}^y\delta_{ij}) \end{aligned}$$

Como los efectos medios de los alelos y los desvíos de dominancia son independientes, entonces:

$$\mathcal{E}({}^x\alpha_i {}^y\delta_{ij}) = \mathcal{E} ({}^x\alpha_j {}^y\delta_{ij}) = \mathcal{E} ({}^x\delta_{ij} {}^y\alpha_i) = \mathcal{E}({}^x\delta_{ij} {}^y\alpha_j) = 0$$

Así como la variancia genotípica se particiona en componentes causales, lo mismo ocurre con la covariancia genotípica. De esta forma podremos descomponerla en covariancia genética o aditiva, covariancia de dominancia y covariancia epistática cuando se consideran dos o más loci.

En ausencia de epistasis y asumiendo que X e Y no son individuos endogámicos, dentro de la covariancia genotípica podemos distinguir dos componentes, la covariancia aditiva o genética y la covariancia de dominancia:

$$\text{Cov}_{G(X,Y)} = \text{Cov}_{AXY} + \text{Cov}_{DXY} = \text{Cov}_G ({}^x\alpha_i + {}^x\alpha_j, {}^y\alpha_i + {}^y\alpha_j) + \text{Cov}_G ({}^x\delta_{ij}, {}^y\delta_{ij})$$

Los $\mathcal{E}({}^x\alpha_i {}^y\alpha_i)$, $\mathcal{E}({}^x\alpha_i {}^y\alpha_j)$, $\mathcal{E}({}^x\alpha_j {}^y\alpha_i)$, $\mathcal{E}({}^x\alpha_j {}^y\alpha_j)$ son iguales a cero cuando X e Y no están emparentados y sus respectivos coeficientes de coancestría son, por lo tanto, iguales a cero. Así $\mathcal{E}({}^x\alpha_i {}^y\alpha_i) = \mathcal{E}({}^x\alpha_i) \mathcal{E}({}^y\alpha_i) = 0$.

Entonces si los individuos X e Y no están emparentados resultará:

$$\text{Cov}_{G(X,Y)} = \text{Cov}_G ({}^x\alpha_i + {}^x\alpha_j + {}^x\delta_{ij}, {}^y\alpha_i + {}^y\alpha_j + {}^y\delta_{ij}) = 0$$

Si en cambio, X e Y estuvieran emparentados por compartir al menos un ancestro en común, la covariancia genotípica entre X e Y será diferente:

$$\text{Cov}_{G(X,Y)} = \text{Cov}_G ({}^x\alpha_i + {}^x\alpha_j + {}^x\delta_{ij}, {}^y\alpha_i + {}^y\alpha_j + {}^y\delta_{ij}) \neq 0$$

En tal caso,

$$\begin{aligned} \text{Cov}_{G(X,Y)} &= \mathcal{E}({}^x\alpha_i + {}^x\alpha_j + {}^x\delta_{ij}) ({}^y\alpha_i + {}^y\alpha_j + {}^y\delta_{ij}) \\ &= \mathcal{E}({}^x\alpha_i {}^y\alpha_i) + \mathcal{E}({}^x\alpha_i {}^y\alpha_j) + \mathcal{E} ({}^x\alpha_j {}^y\alpha_i) + \mathcal{E}({}^x\alpha_j {}^y\alpha_j) + \mathcal{E}({}^x\delta_{ij} {}^y\delta_{ij}), \quad [7.1] \end{aligned}$$

Observemos que la covariancia genotípica posee cuatro términos que involucran a los productos de los efectos medios de los alelos tomados al azar de los individuos X e Y y un término referido a los productos de los desvíos de dominancia. Veamos con más detenimiento cada uno de los términos del miembro de la derecha de la igualdad.

$$\mathcal{E}({}^x\alpha_i {}^y\alpha_i) = f_{XY} \cdot \mathcal{E}({}^y\alpha_i)$$

$$\mathcal{E}(^x\alpha_i \ ^y\alpha_j) = f_{XY} \cdot \mathcal{E}(^y\alpha_j)$$

$$\mathcal{E}(^x\alpha_j \ ^y\alpha_i) = f_{XY} \cdot \mathcal{E}(^y\alpha_i)$$

$$\mathcal{E}(^x\alpha_j \ ^y\alpha_j) = f_{XY} \cdot \mathcal{E}(^y\alpha_j)$$

donde f_{XY} es la probabilidad de que X e Y sean idénticos por descendencia, por definición es el coeficiente de coancestría.

Por lo tanto recordando [7.1], tendremos que

$$\begin{aligned} \text{Cov}_{G(X,Y)} &= \mathcal{E}(^x\alpha_i \ ^y\alpha_i) + \mathcal{E}(^x\alpha_i \ ^y\alpha_j) + \mathcal{E}(^x\alpha_j \ ^y\alpha_i) + \mathcal{E}(^x\alpha_j \ ^y\alpha_j) + \mathcal{E}(^x\delta_{ij} \ ^y\delta_{ij}) \\ &= 4 f_{XY} \mathcal{E}(\alpha^2) + \mathcal{E}(^x\delta_{ij} \ ^y\delta_{ij}), \quad [7.2] \end{aligned}$$

Expresando en términos de variancia aditiva, recordemos que:

$$\sigma_A^2 = \mathcal{E}[(\alpha_i + \alpha_j)^2] = \mathcal{E}(\alpha_i^2) + \mathcal{E}(\alpha_j^2) + 2\mathcal{E}(\alpha_i\alpha_j)$$

Si X no es endogámico, entonces:

$$\begin{aligned} \sigma_A^2 &= \mathcal{E}(\alpha_i^2) + \mathcal{E}(\alpha_j^2) + 2(0) = 2\mathcal{E}(\alpha^2) \\ \mathcal{E}(\alpha^2) &= 1/2 \sigma_A^2, \quad [7.3] \end{aligned}$$

Reemplazando en [7.2]:

$$\text{Cov}_{G(X,Y)} = 4f_{XY} (1/2)\sigma_A^2 + \mathcal{E}(^x\delta_{ij} \ ^y\delta_{ij}) = 2f_{XY} \sigma_A^2 + \mathcal{E}(^x\delta_{ij} \ ^y\delta_{ij})$$

Nos resta determinar el componente $\mathcal{E}(^x\delta_{ij} \ ^y\delta_{ij})$ de la covariancia genotípica.

$$\mathcal{E}(^x\delta_{ij} \ ^y\delta_{ij}) = \mathcal{E}(^x\delta_{ij}) \mathcal{E}(^y\delta_{ij}) = [P(X_i \equiv Y_i, X_j \equiv Y_j) + P(X_i \equiv Y_j, X_j \equiv Y_i)] \mathcal{E}(\delta_{ij}^2) = \Delta_{XY} \sigma_D^2$$

Reemplazando llegamos a la expresión:

$$\text{Cov}_{G(X,Y)} = 2f_{XY} \sigma_A^2 + \Delta_{XY} \sigma_D^2, \quad [7.4]$$

Las medidas de coeficientes de coancestría y de parentesco sirven para cuantificar la covariancia aditiva y el coeficiente de dominancia para cuantificar la covariancia de dominancia entre los individuos X e Y.

Alternativamente podemos llegar al mismo resultado respecto de la covariancia aditiva y la covariancia de dominancia siguiendo el razonamiento que se presenta más adelante. Si los parientes son endocriados, ocurre que los términos $\mathcal{E}(^x\alpha_i \ ^y\delta_{ij})$, $\mathcal{E}(^x\alpha_j \ ^y\delta_{ij})$, $(^x\delta_{ij} \ ^y\alpha_i)$ y $(^x\delta_{ij} \ ^y\alpha_j)$ pueden ser diferentes de cero y en consecuencia la partición de la covariancia genotípica en covariancia aditiva y covariancia de dominancia no es factible, a menos que las frecuencias génicas sean iguales a 0,5.

7.1.1. Covariancia aditiva

Consideremos dos individuos, X e Y, hijos del cruzamiento AxB y CxD, respectivamente, tal como se presenta en el Cuadro 7.1.

1	A (a ₁ a ₂)	B (b ₁ b ₂)	C (c ₁ c ₂)	D (d ₁ d ₂)
2	X (x ₁ x ₂)		Y (y ₁ y ₂)	

Cuadro 7.1.
Esquema de relación genealógica de dos individuos producto del cruzamiento entre individuos de la generación parental.

La covariancia aditiva surge de considerar las posibles situaciones de identidad por descendencia entre los alelos de X e Y.

Es decir, que tenemos que considerar las siguientes alternativas de identidad por descendencia:

$$X1 \equiv Y1; X2 \equiv Y2; X1 \equiv Y2; \text{ y } X2 \equiv Y1;$$

Cada una de estas alternativas tiene una probabilidad promedio de ocurrencia que es el coeficiente de coancestría entre X e Y (f_{XY}).

La covariancia aditiva entre X e Y para el locus en consideración resultará de sumar los aportes de los valores aditivos de cada una de aquellas cuatro alternativas de identidad por descendencia, ponderadas por sus respectivas probabilidades, en este caso su coeficiente de coancestría.

Las cuatro alternativas poseen la misma probabilidad, por lo que:

$$Cov_{A(X,Y)} = 4f_{XY} \cdot \frac{1}{2} V(\alpha_i) = 2f_{XY} V_A$$

7.1.2. Covariancia de dominancia

Asumiendo dos individuos, X e Y, hijos producto del cruzamiento AxB y CxD, respectivamente, tal como se presenta en el diagrama anterior.

Recordemos que la variancia de dominancia es:

$$Cov_D(x\delta_{ij}, y\delta_{ij}) = V(\delta_{ij}) = V_D$$

La covariancia de dominancia surge de considerar las posibles situaciones de identidad por descendencia entre los dos alelos de X y de Y.

Es decir, que tenemos que considerar las siguientes alternativas:

$$X1 \equiv Y1 \text{ y } X2 \equiv Y2; X1 \equiv Y2; \text{ y } X2 \equiv Y1;$$

Ambas alternativas tienen una probabilidad de ocurrencia expresable en términos de producto de los coeficientes de coancestría entre los padres de X e Y:

$$P(X1 \equiv Y1 \text{ y } X2 \equiv Y2) = f_{AC} \times f_{BD}$$

$$P(X1 \equiv Y2 \text{ y } X2 \equiv Y1) = f_{BC} \times f_{AD}$$

En consecuencia:

$$\text{Cov}_{D(X,Y)} = (f_{AC} \times f_{BD} + f_{BC} \times f_{AD})V_D$$

Recordemos que

$$\Delta_{XY} = (f_{AC} \times f_{BD} + f_{BC} \times f_{AD})$$

$$\text{Por lo tanto } \text{Cov}_{D(X,Y)} = \Delta_{XY} V_D$$

Volviendo a la expresión de la covariancia genotípica entre parientes no endogámicos y en ausencia de epistasis:

$$\text{Cov}_{G(X,Y)} = \text{Cov}_{A(X,Y)} + \text{Cov}_{D(X,Y)} = 2f_{XY} V_A + \Delta_{XY} V_D$$

Falconer emplea otra nomenclatura, según la cual:

$$r = 2f_{XY} \text{ y } u = \Delta_{XY}, \text{ por lo que:}$$

$$\text{Cov}_{G(X,Y)} = rV_A + uV_D \quad [7.5]$$

Los coeficientes que acompañan a las variancias aditivas y de dominancia para determinar las covariancias aditiva y de dominancia entre familias de parientes con antecesor común no endogámico se presentan en la Tabla 7.1.

Familia de parientes	$r = 2f_{XY}$	$u = \Delta_{XY}$		
Progenitor-progenie	$\frac{1}{2}$	V_A	0	V_D
Medios hermanos	$\frac{1}{4}$	V_A	0	V_D
Hermanos completos	$\frac{1}{2}$	V_A	$\frac{1}{4}$	V_D
Tío-sobrino	$\frac{1}{4}$	V_A	0	V_D
$S_{0:1}$	1*	V_A	$\frac{1}{4}$	V_D

Tabla 7.1. Coeficientes de ponderación de la variancia aditiva y de la variancia de dominancia entre diferentes tipos de parientes.

En el caso de que el antecesor común sea endocriado, en el cálculo de la covariancia debe tenerse en cuenta esta circunstancia y las expresiones de r y u son diferentes (Tabla 7.2).

Familia de parientes	$r = 2f_{XY}$		$u = \Delta_{XY}$	
Medios hermanos	$(1+F_B)/4$	V_A	0	V_D
Hermanos completos	$(2 + F_A + F_B)/4$	V_A	$[(1+F_A)(1+F_B)]/4$	V_D

Tabla 7.2.

Coefficientes de ponderación de la variancia aditiva y de la variancia de dominancia entre familias de medios hermanos y hermanos completos. Los subíndices A y B se refieren a los progenitores de los hermanos completos y B al padre común de los medios hermanos (Adaptado de Bernardo, 2010).

7.1.3. Covariancia genotípica de medios hermanos

Consideremos el pedigrí que se presenta en el Cuadro 7.2.

1	A(a ₁ a ₂)	B (b ₁ b ₂)	B (b ₁ b ₂)	C(c ₁ c ₂)	Cuadro 7.2. Esquema de genealogía y constitución alélica de dos medios hermanos descendientes de un progenitor en común (B).
2	X (x ₁ x ₂)		Y (y ₁ y ₂)		
3	Z				

A partir de la fórmula general, calculando los coeficientes de coascendencia (f_{XY}), de Kempthorne (r), y de dominancia (Δ_{XY} ; u) podemos estimar cuál es la covariancia para cada tipo de parentesco. También puede estimarse f_{XY} a partir del cálculo del coeficiente de endogamia del descendiente Z (F_Z).

De tal manera, aplicando la fórmula general y según lo visto en el capítulo anterior, por la regla [6.1]:

$$f_{XY} = (1/4)[f_{AB} + f_{AC} + f_{BB} + f_{BC}]$$

Siendo que A, B y C no están emparentados, aplicando la regla [6.3] y asumiendo que B no es endogámico:

$$f_{XY} = (1/4)[0 + 0 + (1/2)(1 + F_B) + 0] = (1/4) [0 + 0 + (1/2)(1 + 0) + 0] = 1/8$$

$$2f_{XY} = r = 1/4$$

$$\Delta_{XY} = f_{AB} \cdot f_{BC} + f_{AC} \cdot f_{BB} = (0) \cdot (0) + (0) \cdot [(1/2)(1 + F_B)] = 0$$

Por lo tanto:

$$\text{Cov}_{G(\text{MH})} = (1/4)V_A + 0V_D = (1/4)V_A$$

7.1.4. Covariancia genotípica en sistemas de endocría aplicados a una población de referencia no endocriada

En ocasión del desarrollo de líneas endocriadas se efectúa selección entre y dentro de familias de manera que puede generarse un entramado de familias emparentadas que comparten iguales o diferentes niveles de endocría. Cuando la población de referencia no es endogámica y las frecuencias génicas son $p = q = 0,5$ (como en el caso de una población biparental resultante del cruzamiento entre dos padres homocigotas), la covariancia genotípica entre familias endocriadas de cada generación puede estimarse a partir de los componentes de variancia genotípica de la población de referencia (Hallauer y Miranda Filho, 1988):

$$\text{Cov}_{G(\text{Si})} = \theta_1 V_A + \theta_2 V_D = (1+F_t) V_A + [(1+F_t)/(1-F_t)](1-F_g)^2 V_D$$

F_t y F_g son los coeficientes de endocría del último padre común en la cadena de descendencia y el coeficiente de endocría de los parientes que están siendo considerados (la i -ésima filial de autofecundación), respectivamente.

Por lo tanto, la covariancia genotípica entre familias de diferentes generaciones de autofecundación ($S_1, S_2, S_3, \text{RILs}$) a partir de una misma planta S_0 en cada cohorte (cada generación surge de la autofecundación de los individuos de la generación previa, sin selección de plantas individuales y tomando el conjunto (bulk) de semillas) será:

$$\begin{aligned} \text{Cov}_{G(S_0:1)} &= (1+F_t) V_A + [(1+F_t)/(1-F_t)](1-F_g)^2 V_D \\ &= (1+0) V_A + [(1+0)/(1-0)](1-0,5)^2 V_D \\ &= V_A + (1/4) V_D \end{aligned}$$

$$\begin{aligned} \text{Cov}_{G(S_0:2)} &= (1+F_t)V_A + [(1+F_t)/(1-F_t)](1-F_g)^2 V_D \\ &= (1+0)V_A + [(1+0)/(1-0)](1-0,75)^2 V_D \\ &= V_A + 1/16 V_D \end{aligned}$$

$$\begin{aligned} \text{Cov}_{G(S_0:3)} &= (1+F_t)V_A + [(1+F_t)/(1-F_t)](1-F_g)^2 V_D \\ &= (1+0)V_A + [(1+0)/(1-0)](1-0,875)^2 V_D \\ &= V_A + (1/64)V_D \end{aligned}$$

$$\text{Cov}_{G(S_0:n)} = (1+F_t)V_A + [(1+F_t)/(1-F_t)](1-F_g)^2 V_D$$

$$\begin{aligned}
 &= (1 + 0)V_A + [(1 + 0)/(1 - 0)](1 - 1)2 V_D \\
 &= V_A
 \end{aligned}$$

Las covariancias genotípicas entre familias de diferentes generaciones de autofecundación (S_1, S_2, S_3, S_n) a partir de una planta con endocría previa seleccionada en la generación anterior, en cada cohorte serán:

$$\begin{aligned}
 \text{Cov}_{G(S_{0:1})} &= (1 + F_t)V_A + [(1 + F_t)/(1 - F_t)](1 - F_g)^2V_D \\
 &= (1 + 0)V_A + [(1 + 0)/(1 - 0)](1 - 0,5)^2V_D \\
 &= V_A + (1/4)V_D
 \end{aligned}$$

$$\begin{aligned}
 \text{Cov}_{G(S_{1:2})} &= (1 + F_t)V_A + [(1 + F_t)/(1 - F_t)](1 - F_g)^2V_D \\
 &= (1 + 0,5)V_A + [(1 + 0,5)/(1 - 0,5)](1 - 0,75)^2V_D \\
 &= 1,5 V_A + (3/16) V_D
 \end{aligned}$$

$$\begin{aligned}
 \text{Cov}_{G(S_{2:3})} &= (1 + F_t)V_A + [(1 + F_t)/(1 - F_t)](1 - F_g)^2V_D \\
 &= (1 + 0,75)V_A + [(1 + 0,75)/(1 - 0,75)](1 - 0,875)^2V_D \\
 &= 1,75 V_A + (7/64)V_D
 \end{aligned}$$

$$\begin{aligned}
 \text{Cov}_{G(S_{n-1:n})} &= (1 + F_t)V_A + [(1 + F_t)/(1 - F_t)](1 - F_g)^2V_D \\
 &= (1 + 1)V_A + [(1 + 0)/(1 - 0)](1 - 1)^2 V_D \\
 &= 2V_A
 \end{aligned}$$

La aplicación de la fórmula general al esquema de endocría y selección que se presenta a continuación permite calcular las siguientes covariancias entre parientes, aun pertenecientes a diferentes generaciones, a condición de que no existan componentes de epistasis que involucren términos aditivos x dominantes ni dominantes x dominantes.

Así por ejemplo, la covariancia entre parientes pertenecientes a diferentes generaciones de endocría (g y g') será:

$$\text{Cov}_{G(SgSg')} = \theta_1 V_A + \theta_2 V_D = (1 + F_t)V_A + [(1 + F_t)/(1 - F_t)](1 - F_g)(1 - F_{g'})V_D$$

Por ejemplo, la covariancia entre familias $S_{1:2}$ y $S_{1:3}$ será:

$$\begin{aligned}
 \text{Cov}_{G(SgSg')} &= \theta_1 V_A + \theta_2 V_D \\
 &= (1 + 0,5)V_A + [(1 + 0,5)/(1 - 0,5)](1 - 0,75)(1 - 0,875)V_D \\
 &= (1,5)V_A + (3/32)V_D
 \end{aligned}$$

y la covariancia entre familias $S_{1:2}$ y $S_{1:5}$ será:

$$\begin{aligned} \text{Cov}_{G(SgSg)} &= \theta_1 V_A + \theta_2 V_D \\ &= (1 + 0,5)V_A + [(1 + 0,5)/(1 - 0,5)](1 - 0,75)(1 - 0,96875)V_D \\ &= (1,5)V_A + (3/128)V_D \end{aligned}$$

La existencia de interacciones epistáticas agrega otro componente causal a la covariancia genotípica. De manera análoga a los componentes de variancia epistática, la covariancia epistática puede descomponerse en función del tipo de acción génica considerada y del número de loci involucrados. Por ejemplo, la contribución de la covariancia epistática aditiva x aditiva al parecido entre parientes implica que un alelo tomado al azar del individuo X y otro del Y en un locus R sean idénticos por descendencia y que lo mismo ocurra en otro locus S. La probabilidad de que ambos sucesos ocurran simultáneamente será:

$$\begin{aligned} &P[(x1 \equiv y1) \text{ o } (x2 \equiv y1) \text{ o } (x1 \equiv y2) \text{ o } (x2 \equiv y2)] \text{ y} \\ &P[(w1 \equiv z1) \text{ o } (w2 \equiv z1) \text{ o } (w1 \equiv z2) \text{ o } (w2 \equiv z2)] \\ &P = (2f_{XY}) \times (2f_{XY}) = 4f_{XY}^2 \end{aligned}$$

Por lo tanto:

$$\text{Cov}(\alpha_i, \alpha_k) = 4f_{xy}^2 V(\alpha_i \alpha_k) = (2f_{XY})^2 V_{AA}$$

La covariancia genotípica entre dos individuos emparentados, X e Y, considerando los términos de covariancia epistática entre dos loci será:

$$\begin{aligned} \text{Cov}_{G(X,Y)} &= \text{Cov}(\alpha) + \text{Cov}(\delta) + \text{Cov}(\alpha\alpha) + \text{Cov}(\alpha\delta) + \text{Cov}(\delta\delta) = \\ &= 2f_{XY} V_A + \Delta_{XY} V_D + (2f_{XY})^2 V_{AA} + 2f_{XY} \Delta_{XY} V_{AD} + \Delta_{XY}^2 V_{DD} \end{aligned}$$

7.1.5. Modelo general de cómputo de la covariancia genotípica entre parientes pertenecientes a una población de referencia no endogámica

Kempthorne (1954) presentó una fórmula general para determinar la covariancia entre parientes pertenecientes a poblaciones no endogámicas. Cada componente de la covariancia genotípica resulta ponderado por un coeficiente C, que se calcula a partir de la siguiente expresión:

$$C = [(\phi + \phi')/2]^a (\phi\phi')^d$$

En esta expresión, a es el número de subíndices de aditividad (A) que figuran en el componente de variancia genotípica y d hace referencia al número de subíndices de dominancia (D) que figuran en el componente de variancia genotípica. De tal manera:

$$\begin{aligned} \text{COV}_{G(X,Y)} = & [(\phi + \phi')/2]^1 (\phi\phi')^0 V_A + [(\phi + \phi')/2]^0 (\phi\phi')^1 V_D + \\ & + [(\phi + \phi')/2]^2 (\phi\phi')^0 V_{AA} + + [(\phi + \phi')/2]^1 (\phi\phi')^1 V_{AD} + \\ & + [(\phi + \phi')/2]^0 (\phi\phi')^2 V_{DD} \end{aligned}$$

Así, por ejemplo, la covariancia genotípica entre familias de hermanos completos (HC) generados en una población no endogámica será:

$$\begin{aligned} \text{COV}_{G(HC)} = & (1/2)^1 (1/4)^0 V_A + (1/2)^0 (1/4)^1 V_D + (1/2)^2 (1/4)^0 V_{AA} + \\ & + (1/2)^1 (1/4)^1 V_{AD} + (1/2)^0 (1/4)^2 V_{DD} + \dots \\ = & (1/2)V_A + (1/4)V_D + (1/4)V_{AA} + (1/8)V_{AD} + (1/16)V_{DD} + \dots \end{aligned}$$

En la Tabla 7.3 se presentan los coeficientes para diferentes tipos de familias empleando los correspondientes coeficientes ϕ y ϕ' .

Parientes	$(\phi+\phi')/2$	$\phi\phi'$
Padre-progenie	$(1+0)/2 = 1/2$	0
Medios hermanos	$(1/2 + 0)/2 = 1/4$	0
Hermanos completos	$(1/2 + 1/2)/2 = 1/2$	$(1/2) \cdot (1/2) = 1/4$


Tabla 7.3. Coeficientes $(\phi + \phi')/2$ y $(\phi\phi')$ (Hallauer y Miranda Filho, 1988) de las variancias aditiva y de dominancia para diferentes tipos de parientes.

7.1.6. Modelo general de cómputo de la covariancia genotípica entre parientes pertenecientes a una población de referencia endogámica

Cockerham (1963) presentó una fórmula general que tiene en cuenta el coeficiente de endogamia en la población de referencia para calcular los coeficientes de parentesco ($r=2f_{XY}$) y de dominancia ($u = \Delta_{XY}$)

$$\text{COV}_{G(X,Y)} = rV_A + uV_D + r^2V_{AA} + ruV_{AD} + u^2V_{DD}$$

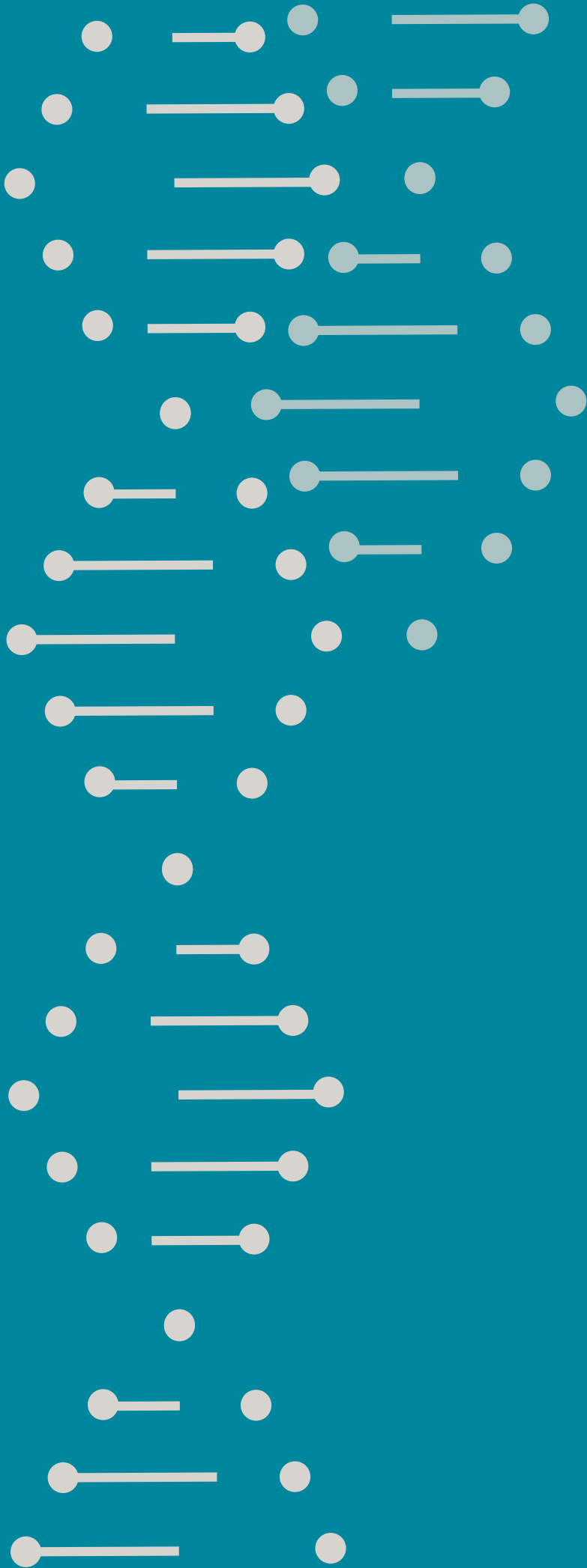
Los coeficientes correspondientes de presentan en la Tabla 7.4 para varios tipos de parientes.



Parientes	$r = 2f_{xy}$	u
Padre-progenie	1/2	0
Medios hermanos	(1+F)/4	0
Hermanos completos	(1+F)/2	[(1+F)/2] ²
Tío-sobrino	(1+F)/4	0
Primos hermanos	(1+F)/8	0

Tabla 7.4.

Coefficientes r y u (Falconer y Mackay, 2001) de las variancias aditiva y de dominancia para diferentes tipos de parientes considerando la endogamia poblacional.





Diseños genéticos

8.1. Concepto y aplicación de los diseños genéticos

El término diseños genéticos se refiere a diseños experimentales cuyos objetivos son estimar, para una población de referencia, las variancias fenotípica, genotípica y sus componentes causales. Ello permite, a su vez, estimar la heredabilidad y posteriormente predecir la respuesta a la selección artificial.

Como todo diseño experimental, un diseño genético requiere de repeticiones y en caso de replicarse en diferentes ambientes, permite estimar también la variancia de interacción genotipo x ambiente y reducir el sesgo en la estimación de la variancia genotípica y sus componentes.

El alcance y la interpretación de los diseños genéticos dependen del cumplimiento de ciertos supuestos:

- La muestra poblacional es representativa de los individuos (por número y elección aleatoria) de una población de referencia que está en equilibrio de Hardy-Weinberg
- Herencia diploide, dos alelos posibles por locus
- Ausencia de efectos recíprocos en los cruzamientos
- Ausencia de epistasis
- Equilibrio de ligamiento

Algunos de los diseños han sido especialmente concebidos para detectar desvíos significativos del cumplimiento de aquellos (desequilibrio de ligamiento, presencia de epistasis, diferencia entre cruzamientos recíprocos).


La complejidad de los diseños genéticos es muy variada y conforme ella aumenta permitirá estimar mayor número de componentes de la variancia genotípica y resultarán así más informativos.

Los diseños genéticos más sencillos son los de un factor. Dentro de esta categoría se incluyen algunos diseños que son rutinariamente empleados en los programas de mejoramiento genético. En la primera parte de este capítulo se presentarán las características de los diseños más simples, su utilidad y limitaciones principales.

8.1.1. Modelos de efectos fijos y aleatorios

Como se mencionó más arriba, los objetivos de los diseños genéticos son lograr estimaciones de las variancias de origen genético, de la heredabilidad y de la respuesta esperable por selección en una población de referencia, por lo cual el modelo aleatorio es generalmente el indicado. En el caso de los experimentos más sencillos y rutinarios en los programas de mejoramiento puede ocurrir que además de estimar la respuesta a la selección, interese determinar los efectos genéticos de los genotipos evaluados considerando que el conjunto de ellos no constituye una muestra de una población de referencia, sino la totalidad de esta. En este caso el objetivo primario es la selección de los genotipos de mejor comportamiento para el carácter de interés y se emplea el modelo de efectos fijos.

Entender cuál es el modelo apropiado en un diseño genético tiene implicancias muy importantes en el análisis estadístico, ya que las esperanzas matemáticas de los cuadrados medios varían según el modelo y consecuentemente los términos de error que deben emplearse en las pruebas de



F de Snedecor para probar la significancia de las variancias (modelo aleatorio) y de los efectos (modelo de efectos fijos) son diferentes.

Para decidir si se trata de un caso de una variable clasificatoria (por ejemplo genotipos, ambientes) de efectos fijos o de efectos aleatorios, podemos orientarnos teniendo en cuenta los siguientes criterios:

Estaremos en el caso de una variable clasificatoria de efectos fijos cuando:

- Todos los niveles de la variable o factor clasificatorio (genotipos) han sido incorporados al experimento, o solo los niveles que resultan de interés. Es cuando, por ejemplo, se evalúa un grupo concreto o específico de genotipos y los resultados se refieren exclusivamente a ellos. Si la evaluación de una serie de genotipos se hiciera en condiciones ambientales adecuadamente controlables a diferentes niveles, y al investigador le interesara conocer el comportamiento de esos genotipos en ese grupo definido y concreto de situaciones ambientales, la variable clasificatoria “ambiente” sería de tipo fijo.
- Los niveles del factor (genotipos) no han sido elegidos al azar. Este concepto esencialmente está relacionado con el anterior. Los genotipos no constituyen una muestra aleatoria de un grupo mayor, sino que su elección ha sido deliberada.
- Tales niveles pueden repetirse en otro experimento y las variaciones que pueden encontrarse son debidas a efectos aleatorios. Los genotipos (cruzamientos híbridos, clones, líneas puras de especies autóгамas, familias intrapoblacionales o interpoblacionales) pueden evaluarse en otro experimento por cuanto es posible mantener su identidad genotípica, aunque los resultados pueden diferir por efectos aleatorios.

Estaremos ante un caso de variable clasificatoria aleatoria:

- Cuando todos los niveles incluidos en el experimento de una variable clasificatoria han sido extraídos al azar de una población de referencia, conceptual o definible. Por ejemplo cuando se originan familias al azar en una población para constituirse en una muestra aleatoria de la población para estimar sus parámetros poblacionales.
- Cuando la repetición del experimento podría resultar en un nuevo grupo de niveles de la variable clasificatoria. Si en el caso ejemplificado más arriba se volvieran a generar familias de la población, es virtualmente imposible obtener los mismos genotipos. En el caso de un experimento que se realiza en una serie de ambientes, definidas como “localidades”, es imposible que esas condiciones ambientales puedan reproducirse al año siguiente.

- Si la variabilidad observada es resultado de la suma de las variabilidades aportadas por la propia población de referencia y el error experimental.

Mayor información sobre los modelos de efectos fijos y efectos aleatorios puede consultarse en el Anexo.

8.2. Consideraciones prácticas

Una eficiente evaluación de genotipos, especialmente cuando su número es relativamente grande como el que se requiere para lograr estimaciones representativas de las variancias poblacionales, tiene como condición lograr un buen control local de la variabilidad ambiental. Cuanto mayor sea el número de genotipos, mayor dimensión tendrá cada repetición del experimento y menor será la probabilidad de lograr condiciones uniformes de suelo, relieve, etc., dentro de cada una de ellas.

Para mitigar el problema, se han propuesto numerosos diseños experimentales, como por ejemplo los de bloques incompletos (*lattices*, *alpha lattices*), análisis estadístico con ajustes por fila y columna, por vecinos más cercanos, etc. Todas estas herramientas ajustan las medias observadas de los genotipos evaluados en función de alguna corrección por efectos ambientales y resultan muy convenientes cuando el propósito principal es lograr confiables estimaciones de medias genotípicas con fines de selección. Sin embargo, por el propósito básico de los diseños genéticos (estimación de variancias) resulta preferible adoptar un diseño experimental que en la medida de lo posible no modifique los datos observados en base a esos ajustes ambientales.

Para lograrlo, una alternativa para mejorar el control de la variabilidad ambiental dentro del sitio experimental es distribuir aleatoriamente a los genotipos para evaluar en grupos o sets, cada uno de ellos evaluado en repeticiones o bloques completos de menores dimensiones.

8.3. Evaluación de genotipos en experimentos con repeticiones

Cuando se trata de disponer de estimaciones de variancias genéticas poblacionales y heredabilidad para caracteres sobre los que hay abundante información previa, o cuando se dispone de experiencia en el tema y de estimaciones sobre otras poblaciones similares a las de referencia, puede no ser necesario utilizar diseños genéticos que impliquen el desarrollo de

progenies o familias diferentes a las que rutinariamente se evalúan en un programa de mejoramiento. En estas situaciones, la evaluación de este tipo de material genético puede resultar suficientemente informativa. En el caso de especies de reproducción agámica puede evaluarse un conjunto de clones en ensayos con repeticiones en uno o más ambientes. Cuando se trata de especies autógamias, pueden evaluarse líneas en desarrollo y en alógamias evaluarse líneas endocriadas per se o en cruzamientos de prueba con un probador común a todas ellas.

El esquema general del análisis de variancia (ANdeVA) cuando se evalúan los genotipos en experimentos en un único ambiente y con r repeticiones se presenta en la Tabla 8.1.

Fuentes de variación	Grados de libertad	Cuadrado medio	ϵ (CM)	F
Repeticiones	r-1			
Genotipos	g-1	CM _g	$\sigma_e^2 + r\sigma_g^2$	CM _g /CM _{EE}
Error experimental	(r-1)(g-1)	CM _{EE}	σ_e^2	
Total	rg -1			

Tabla 8.1.

Esquema de un análisis de variancia y esperanza de los cuadrados medios donde se evalúan g genotipos en un experimento con r repeticiones.

Obsérvese que a los fines de estimar variancias debemos considerar a “Genotipos” como una variable de efecto aleatorio.

Puede estimarse la variancia genotípica (V_g) a partir de:

$$(CM_g - CM_{EE}) / r = V_g$$

Dependerá de la naturaleza de los genotipos que se están evaluando el tipo de heredabilidad factible de estimar. Si se evalúan clones, puede lograrse una estimación de la heredabilidad en sentido amplio:

$$H = V_g / (V_g + V_{EE}) \text{ (sobre la base de plantas individuales)}$$

y

$$H = V_g / (V_g + V_{EE}/r) \text{ (sobre la base de media de clones a través de r repeticiones).}$$

La heredabilidad estimada lo será en sentido amplio debido a que tanto la variancia aditiva como la no aditiva pueden estar contribuyendo a la variancia genotípica entre clones.

Si se evaluaran líneas recombinantes estabilizadas o haploides duplicados, sabemos que la variancia entre ellas estima el doble de la variancia aditiva de la población de referencia (ver Capítulo 5). En tal caso deberemos tener presente que:

$$(CM_g - CM_{EE})/r = V_g = 2V_A ,$$

Donde V_A es la variancia genética o aditiva poblacional. Por lo tanto podemos obtener una estimación de la heredabilidad en sentido estricto:

$$h^2 = (1/2)V_g / [(1/2)V_g + V_{EE}] , \text{ (sobre la base de plantas individuales de la población a partir de la cual se generaron las líneas homocigotas),}$$

y

$$h^2 = V_g / (V_g + V_{EE}/r) \text{ (sobre la base de promedio de líneas completamente endocriadas evaluadas en r repeticiones).}$$

Si se evaluaran familias de hermanos completos, sabemos que la variancia entre ellas estima un medio de la variancia aditiva y un cuarto de la variancia de dominancia de la población de referencia (en ausencia de epistasis). En tal caso deberemos tener presente que:

$$(CM_g - CM_{EE})/r = V_g = (1/2)V_A + (1/4)V_D ,$$

Donde V_A y V_D son las variancias aditiva y de dominancia poblacionales, respectivamente. Por lo tanto podemos obtener una estimación grosera (por el sesgo debido a la sobreestimación que implica que V_D pueda ser distinta de cero) de la heredabilidad en sentido estricto:

$$h^2 = 2V_g / (2V_g + V_{EE}) \text{ (sobre la base de planta individual),}$$

y

$$h^2 = V_g / (V_g + V_{EE}/r) \text{ (sobre la base de promedio de familias de hermanos completos a través de repeticiones).}$$

Es usual que estos experimentos de evaluación de genotipos se repliquen en varios ambientes. Ello permite estimar la variancia de interacción genoti-

po x ambiente, que de otra manera, cuando se evalúa en un único ambiente, aparece confundida dentro de V_g .

El esquema general del ANdeVA cuando se evalúan los genotipos en experimentos en a ambientes y con r repeticiones se presenta en la Tabla 8.2.

Fuentes de variación	Grados libertad	Cuadrado medio	ϵ (CM)	F
Ambientes	a-1			
Repeticiones (ambientes)	a(r-1)			
Genotipos	g-1	CM _g	$\sigma_e^2 + r\sigma_{ga}^2 + ar\sigma_g^2$	CM _g /CM _{ga}
Genotipo x Ambiente	(a-1)(g-1)	CM _{ga}	$\sigma_e^2 + r\sigma_{ga}^2$	CM _{ga} /CM _{EE}
Error experimental	a(r-1)(g-1)	CM_{EE}	σ_e^2	
Total	arg -1			

Tabla 8.2.

Esquema de un análisis de variancia y esperanza de los cuadrados medios de la evaluación de g genotipos en experimentos con r repeticiones en a ambientes.

A partir de la resolución de este sistema de tres ecuaciones (igualación de los cuadrados medios con las esperanzas de los cuadrados medios respectivos) podremos obtener las estimaciones de las variancias del error experimental (V_{EE}), de la interacción genotipo x ambiente (V_{GA}) y genotípica (V_G) del siguiente modo:

$$CM_{EE} = V_E$$

$$(CM_{ga} - CM_{EE})/r = V_{GA}$$

$$(CM_g - CM_{ga})/ar = V_G$$

Con este diseño de un factor y según los genotipos que se evalúen, se pueden obtener las siguientes estimaciones de heredabilidad en sentido amplio o estricto, sobre la base de planta individual y de medias de familias a partir de r repeticiones en a ambientes (Tabla 8.3).

Obsérvese que la evaluación en más de un ambiente permite incrementar la heredabilidad sobre la base de promedio de genotipos, al disminuir la contribución de las variancias de interacción genotipo x ambiente y del error experimental a la variancia fenotípica, en función del número de ambientes y del producto del número de ambientes y de repeticiones, respectivamente.

	Heredabilidad en sentido amplio		Heredabilidad en sentido estricto	
	Entre plantas individuales de la población de referencia	Sobre la base de medias de genotipos	Entre plantas individuales de la población de referencia	Sobre la base de medias de genotipos
Clones	$V_g / (V_g + V_{ga} + V_{EE})$	$V_g / (V_g + V_{ga} / a + V_{EE} / ar)$	--	--
Medios hermanos	$4V_g / (4V_g + 4V_{ga} + V_{EE})^{\S}$	$V_g / (V_g + V_{ga} / a + V_{EE} / ar)^{\S}$	$4V_g / (4V_g + 4V_{ga} + V_{EE})$	$V_g / (V_g + V_{ga} / a + V_{EE} / ar) ;$ $1/4 V_A / (1/4 V_A + 1/4 V_{Aa} / a + V_{EE} / ar)$
Líneas con F=1	$1/2 V_g / (1/2 V_g + 1/2 V_{ga} + V_{EE})^{\S}$	$V_g / (V_g + V_{ga} / a + V_{EE} / ar)^{\S}$	$1/2 V_g / (1/2 V_g + 1/2 V_{ga} + V_{EE})$	$V_g / (V_g + V_{ga} / a + V_{EE} / ar) ;$ $2V_A / (2V_A + 2V_{Aa} / a + V_{EE} / ar)$
Hermanos completos	$2V_g / (2V_g + 2V_{ga} + V_{EE})^{\ddagger}$	$V_g / (V_g + V_{ga} / a + V_{EE} / ar)^{\ddagger}$	$2V_g / (2V_g + 2V_{ga} + V_{EE})^{\ddagger}$	$V_g / (V_g + V_{ga} / a + V_{EE} / ar)^{\ddagger} ;$ $[1/2 V_A / (1/2 V_A + 1/4 V_D)]$ $+ (1/2 V_{Aa} + 1/4 V_{Da}) / a$ $+ V_{EE} / ar]$

[§] Expresión resulta sesgada, por cuanto no hay contribución de la variancia de dominancia poblacional.

[†] Expresión resulta sesgada ya que sólo tiene en cuenta el 50 % de la variancia de dominancia poblacional.

[‡] Expresión sobrestimada por cuanto solo tiene en cuanto el 25 % de la contribución de la variancia de dominancia a la variancia fenotípica.

Tabla 8.3.

Heredabilidad en sentido amplio y estricto para diferentes clases de genotipos evaluados en experimentos con repeticiones y replicados en varios ambientes. V_g es el estimador de σ_g^2 , V_{ga} es el estimador de σ_{ga}^2 y V_{EE} es el estimador de σ_c^2 .

8.4. Diseño de progenies biparentales o de hermanos completos

Este diseño fue inicialmente propuesto por Mather (1949) y el material genético para evaluar son cruzamientos de pares de individuos tomados al azar de una población de referencia y en número suficiente para constituir una muestra representativa de la población. Los cruzamientos constituyen familias de hermanos completos. Considerando que n es el número de individuos muestreados de la población, se formarán n/2 familias de hermanos completos.

Cuando se toman observaciones en k individuos pertenecientes a un mismo cruzamiento, el esquema del ANdeVA se presenta en la Tabla 8.4.



Fuentes de variación	Grados de libertad	Cuadrados medios (CM)	ε (CM)
Entre cruzamientos	(n/2) -1	CM _{EC}	σ _{DC} ² + kσ _{EC} ²
Dentro de cruzamientos	(k-1)(n/2)	CM _{DC}	σ _{DC} ²
Total	k(n/2)-1		

Tabla 8.4.

Esquema del análisis de variancia y esperanza de los cuadrados medios para un experimento donde se evalúan progenies de medios hermanos en un experimento con repeticiones, tomando observaciones de k plantas por parcela.

Cuando la prueba de F (CM_{EC}/CM_{DC}) resulta significativa, ello implica que las diferencias entre familias son mayores que las diferencias o variabilidad dentro de ellas. Recordando la composición de las variancias genotípicas entre y dentro de familias de hermanos completos (ver Capítulos 5 y 7):

$$\sigma_{EC}^2 = (\frac{1}{2})\sigma_A^2 + (\frac{1}{4})\sigma_D^2$$

$$\sigma_{DC}^2 = \sigma_w^2 = \sigma_{we}^2 + \{\sigma_G^2 - [(\frac{1}{2})\sigma_A^2 + (\frac{1}{4})\sigma_D^2]\},$$

El coeficiente de correlación intraclase (r_I) puede estimarse como:

$$r_I = \sigma_{EC}^2 / (\sigma_{DC}^2 + \sigma_{EC}^2)$$

$$r_I = [(\frac{1}{2})V_A^2 + (\frac{1}{4})V_D^2] / (\sigma_{we}^2 + \sigma_A^2 + \sigma_D^2)$$

En este análisis de variancia la variancia del error está confundida dentro de los cuadrados medios de ambas fuentes de variación, por lo cual no es posible determinar la significancia de los componentes de la variancia genotípica. Para poder hacerlo es necesario disponer de repeticiones. En este caso, la clave del ANdeVA se presenta en la Tabla 8.5.

Fuentes de variación	Grados de libertad	Cuadrado medio (CM)	ε (CM)
Repeticiones	r-1		
Entre cruzamientos	(n/2) -1	CM _{EC}	σ _w ² + kσ _p ² + rkσ _{EC} ² = σ _w ² + kσ _p ² + rkCov _{G(HC)} = σ _w ² + kσ _p ² + rk(½σ _A ² + ¼σ _D ²)
Error experimental	(r-1)[(n/2)-1]	CM _{EE}	σ _w ² + kσ _p ²
Total	r(n/2) -1		
Dentro de cruzamientos	(k-1)r(n/2)	CM _{DC}	σ _w ² = σ _{we} ² + [σ _G ² - Cov _{G(HC)}] = = σ _{we} ² + [σ _G ² - (½σ _A ² + ¼σ _D ²)]

Tabla 8.5.

Esquema básico del análisis de variancia y esperanza de los cuadrados medios para un experimento de g progenies biparentales, con r repeticiones y tomando observaciones sobre k individuos dentro de cada progenie.

Las esperanzas de los cuadrados medios serán:

$$\varepsilon (CM_{DC}) = \sigma_w^2 = \sigma_{we}^2 + \sigma_{wg}^2 = \sigma_{we}^2 + [\sigma_G^2 - (1/2\sigma_A^2 + 1/4\sigma_D^2)] = \sigma_{we}^2 + 1/2\sigma_A^2 + 3/4\sigma_D^2$$

$$\varepsilon (CM_{EC}) = \sigma_{EC}^2 = \sigma_w^2 + k\sigma_p^2 + kr\sigma_{EC}^2 = \sigma_{we}^2 + (1/2\sigma_A^2 + 3/4\sigma_D^2) + k\sigma_p^2 + rk(1/2\sigma_A^2 + 1/4\sigma_D^2)$$

En la columna de la derecha se reemplazaron las expresiones de la covariancia entre hermanos completos por sus componentes causales de variancia aditiva y de dominancia.

Obsérvese que lo dispuesto en el diseño es un sistema de tres ecuaciones con cuatro incógnitas (σ_{we}^2 , σ_e^2 , σ_A^2 y σ_D^2), por lo que no es posible estimarlas a todas por cuanto no hay suficientes grados de libertad.

Si asumiéramos que no existe variancia de dominancia, el sistema quedaría reducido a tres parámetros incógnitas y tres ecuaciones:

$$CM_{EC} = \sigma_{we}^2 + k\sigma_p^2 + rk(1/2)\sigma_A^2 = \sigma_{we}^2 + (1/2)\sigma_A^2 + k\sigma_p^2 + rk(1/2)\sigma_A^2$$

$$CM_{EE} = \sigma_{we}^2 + k\sigma_p^2 = \sigma_{we}^2 + (1/2)\sigma_A^2 + k\sigma_p^2$$

$$CM_{DF} = \sigma_{we}^2 + (1/2)\sigma_A^2$$

Por lo tanto:

$$\sigma_e^2 = (CM_{EE} - CM_{DC})/k$$

$$\sigma_A^2 = 2(CM_{EC} - CM_{EE})/kr$$

$$\sigma_{we}^2 = CM_{DC} - [(CM_{EC} - CM_{EE})/kr]$$

En ausencia de dominancia, entonces, podría obtenerse una expresión de la importancia de la variancia aditiva de la población respecto de la variancia fenotípica. Esta última es la suma de la variancia aditiva y del error dentro y entre unidades experimentales, dando como resultado la heredabilidad del carácter medido sobre la base de plantas individuales:

$$h_i^2 = \sigma_A^2 / (\sigma_A^2 + \sigma_p^2 + \sigma_{we}^2),$$

En caso de existir dominancia, es posible estimar la variancia genotípica (σ_G^2) y la heredabilidad en sentido amplio o grado de determinación genética (GDG) sobre la base de plantas individuales:

$$CM_{EC} = \sigma_{we}^2 + k\sigma_p^2 + rk\sigma_G^2 = \sigma_{we}^2 + (1/2\sigma_A^2 + 3/4\sigma_D^2) + k\sigma_p^2 + rk(1/2\sigma_A^2 + 1/4\sigma_D^2)$$

$$CM_{EE} = \sigma_{we}^2 + k\sigma_p^2 = \sigma_{we}^2 + 1/2\sigma_A^2 + 3/4\sigma_D^2 + k\sigma_p^2$$

$$CM_{DF} = \sigma_{we}^2 + 1/2\sigma_A^2 + 3/4\sigma_D^2$$

$$GDG = H_i = 2\sigma_{EC}^2 / (2\sigma_{EC}^2 + \sigma_p^2 + \sigma_{ed}^2)$$

La expresión de heredabilidad en sentido estricto, basada en medias de familias evaluadas en r repeticiones será:

$$h_i^2 = \sigma_{EC}^2 / [\sigma_{EC}^2 + (\sigma_p^2 + \sigma_{we}^2)/r]$$

Cuando los cruzamientos de hermanos completos se evalúan en más de un ambiente y sin evaluar dentro de familias, el esquema del ANdeVA se presenta en la Tabla 8.6.

Fuentes de variación	Grados de libertad	Cuadrado medio (CM)	ϵ (CM)	F
Ambientes	a-1			
Repeticiones(amb)	a(r-1)			
Entre cruzamientos	(n/2) - 1	CM _{EC}	$\frac{\sigma_e^2 + r\sigma_{ECxA}^2 + ar\sigma_{EC}^2 = \sigma_e^2 + r\sigma_{ECxA}^2 + arCov_{G(HC)} = \sigma_e^2 + r\sigma_{ECxA}^2 + ar(\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2)}$	CM _{EC} /CM _{ECxA}
Entre Cruzam. x Amb.	(a-1)[(n/2)-1]	CM _{ECxA}	$\sigma_e^2 + r\sigma_{ECxA}^2$	CM _{ECxA} /CM _{EE}
Error experimental	a(r-1)[(n/2)-1]	CM _{EE}	σ_e^2	
Total	ar(n/2) - 1			

Tabla 8.6.

Esquema básico del análisis de variancia y esperanza de los cuadrados medios para un experimento de g progenies biparentales, con r repeticiones en a ambientes.

La estimación de la heredabilidad del carácter en la población basada en plantas individuales será:

$$h_i^2 = 2\sigma_{EC}^2 / [2\sigma_{EC}^2 + 2\sigma_{ECxA}^2 + \sigma_e^2] \text{ (asumiendo que no hay dominancia)}$$

La estimación de la heredabilidad basada en medias de familias evaluadas en r repeticiones en a ambientes será:

$$h_i^2 = \sigma_{EC}^2 / [\sigma_{EC}^2 + \sigma_{ECxA}^2/a + \sigma_e^2/ar] \text{ (asumiendo que no hay dominancia)}$$

8.5. Diseño 1 de Carolina del Norte (NC1)

Este diseño fue propuesto por Comstock y Robinson (1948) y utiliza cruzamientos entre individuos de la población de manera de generar progenies de medios hermanos. Tales individuos deben tomarse al azar de la población de referencia y en número suficiente a fin de obtener una muestra representativa de la población para el carácter de interés. La particularidad

del NC1 es que la estructura de tratamientos (genotipos) es de tipo anidado (ver Anexo) ya que los individuos que se emplean como progenitores machos se cruzan con diferentes hembras, es decir, que las hembras están “anidadas dentro de machos”, como se indica en el esquema que sigue. En este caso se cruzan m machos por 5 hembras diferentes cada uno, de modo que el número total de hembras será mh .

El esquema de cruzamiento anidado en el diseño genético NC1 se representa en el Cuadro 8.1, donde M_i y H_j corresponden a los individuos muestreados de la población y empleados como machos y hembras de los cruzamientos.

M_1	M_2	M_3	...	M_m
H_1	H_6	H_{11}	...	$H_{(5m+1)}$
H_2	H_7	H_{12}	...	$H_{(5m+2)}$
H_3	H_8	H_{13}	...	$H_{(5m+3)}$
H_4	H_9	H_{14}	...	$H_{(5m+4)}$
H_5	H_{10}	H_{15}	...	H_{hm}

Cuadro 8.1.

Esquema de estructura anidada del diseño Carolina del Norte 1. M y H hacen referencia a los progenitores masculinos y femeninos.

Cada macho apareado con sus respectivas hembras da lugar a mh familias de hermanos completos (HC). A su vez todas las progenies de hermanos completos que tienen un mismo macho en común dan lugar a m familias de medios hermanos (MH). Entonces el esquema de cruzamiento anidado da lugar a los tipos de progenies que se indican en el Cuadro 8.2.

M_1	M_2	M_3	...	M_m
$HC_{1,1}$	$HC_{6,2}$	$HC_{11,3}$...	$HC_{(5m+1),m}$
$HC_{2,1}$	$HC_{7,2}$	$HC_{12,3}$...	$HC_{(5m+2),m}$
$HC_{3,1}$	$HC_{8,2}$	$HC_{13,3}$...	$HC_{(5m+3),m}$
$HC_{4,1}$	$HC_{9,2}$	$HC_{14,3}$...	$HC_{(5m+4),m}$
$HC_{5,1}$	$HC_{10,2}$	$HC_{15,3}$...	HC_{hm}
MH_1	MH_2	MH_3	...	MH_m

Cuadro 8.2.

Tipos de familias de hermanos completos (HC) y de medios hermanos (MH) generados por el diseño Carolina del Norte 1.

8.5.1. Análisis de NC1 en un ambiente de evaluación

El modelo estadístico es:

$$Y_{ijkl} = \mu + s_i + r(s)_{ij} + m_k + h(m)_{kl} + e_{ijkl}$$

donde:

Y_{ijklm} es la observación correspondiente al i-ésimo grupo o set del macho k-ésimo cruzado con la l-ésima hembra en la j-ésima repetición

μ es la media general;

s_i es el efecto del i-ésimo grupo o set en el que fueron agrupados los genotipos;

m_k es el efecto de k-ésimo macho;

$h(m)_{kl}$ es el efecto de la l-ésima hembra cruzada con el k-ésimo macho y

e_{ijkl} es el error experimental de la observación correspondiente al i-ésimo set, kl-ésimo genotipo en la j-ésima repetición.

El esquema del ANDeVA para un diseño NC1 efectuado distribuyendo aleatoriamente los genotipos en s grupos, en un único ambiente con r repeticiones dentro de grupos se presenta en la Tabla 8.7.

Fuentes de variación	Grados de libertad	Cuadrados medios (CM)	ϵ (CM)	F
Repeticiones (grupos)	s(r-1)			
Grupos	s-1			
Machos (grupos)	s(m-1)	CM_M	$\sigma_e^2 + r\sigma_{H(M)}^2 + rh\sigma_M^2$	$CM_M/CM_{H(M)}$
Hembras (machos)(grupos)	sm(h-1)	$CM_{H(M)}$	$\sigma_e^2 + r\sigma_{H(M)}^2$	$CM_{H(M)}/CM_{EE}$
Error experimental	s(r-1)(mh-1)	CM_{EE}	σ_e^2	
Total	rsmh -1			

Tabla 8.7.

Esquema del análisis de variancia y esperanzas de los cuadrados medios para un diseño NCI, donde los genotipos se distribuyen aleatoriamente en s grupos, en un único ambiente de evaluación con r repeticiones.

En el esquema que presenta la estructura de las familias, vemos que la progenie de cada uno de los machos constituye una familia de medios hermanos, de manera que la esperanza del cuadrado medio de "Machos" resulta ser:

$$\varepsilon (CM_M) = \sigma_e^2 + r\sigma_{H(M)}^2 + rh\sigma_M^2 = \sigma_e^2 + r\sigma_{H(M)}^2 + rhCOV_{MH} = \sigma_e^2 + r\sigma_{H(M)}^2 + rh\frac{1}{4}\sigma_A^2$$

Con el mismo razonamiento podemos entender la naturaleza de la variabilidad entre hembras dentro de macho si tenemos en cuenta que entre ellas hay un factor que contribuye a su diferenciación y es que se trata de familias de hermanos completos. Sin embargo existe otro factor que, en cambio, disminuye la variabilidad entre hembras dentro de macho: todas ellas comparten un macho en común y por lo tanto son medios hermanos. En consecuencia:

$$\begin{aligned} \varepsilon (CM_{H(M)}) &= \sigma_e^2 + r\sigma_{H(M)}^2 = \sigma_e^2 + r(COV_{HC} - COV_{MH}) \\ &= \sigma_e^2 + r[(\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2) - \frac{1}{4}\sigma_A^2] \\ &= \sigma_e^2 + r(\frac{1}{4}\sigma_A^2 + \frac{1}{4}\sigma_D^2) \end{aligned}$$

El sistema de ecuaciones constituido por las esperanzas de los cuadrados medios permite lograr estimaciones de las variancias de origen genético y de la heredabilidad.

Hemos indicado más arriba que la variancia entre machos equivale a la covariancia de medios hermanos, por lo tanto:

$$V_M = (CM_M - CM_{H(M)})/rh = \frac{1}{4} V_A$$

En consecuencia, la variancia aditiva estimada en la población de referencia será:

$$V_A = 4V_M = 4(CM_M - CM_{H(M)})/rh$$

La variancia de hembras dentro de machos puede estimarse como

$$V_{H(M)} = (CM_{H(M)} - CM_{EE})/r = \frac{1}{4}\sigma_A^2 + \frac{1}{4}\sigma_D^2$$

La variancia genotípica total en la población de referencia, asumiendo ausencia de epistasis será entonces

$$V_G = 4V_{H(M)} = V_A + V_D$$

Por lo tanto podemos lograr una estimación de la variancia de dominancia de la siguiente manera

$$V_D = V_G - V_A = 4V_{H(M)} - 4V_M = 4(V_{H(M)} - V_M)$$

La estimación de la variancia del error será

$$V_{EE} = CM_{EE}$$

Con las estimaciones disponibles es posible evaluar las heredabilidades en sentido amplio y estricto sobre la base de plantas individuales en la población de referencia de la siguiente manera

$$H = V_G / (V_G + V_{EE}) = 4V_{H(M)} / (4V_{H(M)} + V_{EE})$$

$$h^2 = V_A / (V_G + V_{EE}) = 4V_M / (4V_{H(M)} + V_{EE})$$

Las heredabilidades en sentido amplio y estricto para las medias de genotipos evaluados en un ambiente con r repeticiones serán (Hallauer y Miranda Filho, 1988).

$$H = V_G / (V_G + V_{EE}/r) = 4V_{H(M)} / (4V_{H(M)} + V_{EE}/r)$$

$$h^2 = V_A / (V_G + V_{EE}/r) = 4V_M / (4V_{H(M)} + V_{EE}/r)$$

8.5.2. Análisis de NC1 en varios ambientes de evaluación

El modelo estadístico es:

$$Y_{ijklm} = \mu + a_j + s(a)_{ij} + r(as)_{ijk} + m_l + h(m)_{lm} + ma_{jl} + h(m)a_{jlm} + e_{ijklm}$$

donde

Y_{ijklm} es la k-ésima observación correspondiente al i-ésimo set o grupo del macho l-ésimo cruzado con la m-ésima hembra en el j-ésimo ambiente;

μ es la media general;

a_j es el efecto del j-ésimo ambiente;

$s(a)_{ij}$ es el efecto del i-ésimo set anidado en el j-ésimo ambiente;

$r(as)_{ijk}$ es el efecto de la k-ésima repetición anidada en el j-ésimo ambiente e i-ésimo set.

m_l es el efecto de l-ésimo macho;

$h(m)_{lm}$ es el efecto de la m-ésima hembra cruzada con el l-ésimo macho;

ma_{jl} es el efecto de interacción del l-ésimo macho en el j-ésimo ambiente;

$h(m)a_{jl}$ es el efecto de interacción de la m-ésima hembra anidada en el l-ésimo macho en el j-ésimo ambiente;

e_{ijklm} es el error experimental asociado al lm-ésimo genotipo en el la k-ésima repetición anidada en el i-ésimo set y j-ésimo ambiente.

El esquema del análisis de variancia para un diseño NC1 distribuyendo los genotipos en s grupos, en a ambientes y con r repeticiones dentro de grupos se presenta en la Tabla 8.8.

Fuentes de variación	Grados de libertad (gl)	Cuadrados medios (CM)	ϵ (CM)	F
Ambientes	a-1			
Repeticiones (Amb.)	a(r-1)			
Grupos (Rep.)(Amb.)	ar(s-1)			
Machos (Grupos)	s(m-1)	CM_M	$\sigma_e^2 + r\sigma_{H(M)xA}^2 + rh\sigma_{MxA}^2 + ar\sigma_{H(M)}^2 + arh\sigma_M^2$	$CM_M / [CMS]^*$
Hembras (Machos)(Grupos)	sm(h-1)	$CM_{H(M)}$	$\sigma_e^2 + r\sigma_{H(M)xA}^2 + ar\sigma_{H(M)}^2$	$CM_{H(M)} / CM_{H(M)xA}$
Machos (Grupos) x Amb.	(m-1)(a-1)	CM_{MxA}	$\sigma_e^2 + r\sigma_{H(M)xA}^2 + rh\sigma_{MxA}^2$	$CM_{MxA} / CM_{H(M)xA}$
Hembras (Machos)(Grupos) x Amb.	m(h-1)(a-1)	$CM_{H(M)xA}$	$\sigma_e^2 + r\sigma_{H(M)xA}^2$	$CM_{H(M)xA} / CM_{EE}$
Error experimental	as(r-1)(mh-1)	CM_{EE}	σ_e^2	
Total	arsmh -1			

* $[CMS] = CM_{MxA} + CM_{H(M)} - CM_{H(M)xA}$

Tabla 8.8.

Esquema básico del análisis de variancia y esperanzas de los cuadrados medios para un diseño NC1 donde los genotipos se agruparon en s grupos, con r repeticiones dentro de grupos y replicados en a ambientes.

En el esquema que presenta la estructura de las familias vemos que la progenie de cada macho constituye una familia de medios hermanos, de manera que la esperanza del cuadrado medio de "Machos" resulta ser:

$$\begin{aligned}
 \epsilon(CM_M) &= \sigma_e^2 + r\sigma_{H(M)xAmb}^2 + rh\sigma_{MxAmb}^2 + ar\sigma_{H(M)}^2 + arh\sigma_M^2 \\
 &= \sigma_e^2 + r\sigma_{H(M)xAmb}^2 + rh\sigma_{MxAmb}^2 + ar\sigma_{H(M)}^2 + arhCOV_{MH} \\
 &= \sigma_e^2 + r\sigma_{H(M)xAmb}^2 + rh\sigma_{MxAmb}^2 + ar\sigma_{H(M)}^2 + arh\frac{1}{4}\sigma_A^2
 \end{aligned}$$

Con el mismo razonamiento podemos entender la naturaleza de la variabilidad entre hembras dentro de macho si tenemos en cuenta que entre ellas hay un factor que contribuye a su diferenciación y es que se trata de familias de hermanos completos. Sin embargo existe un factor que disminuye la variabilidad entre hembras dentro de macho: todas ellas al compartir un macho son medios hermanos. En consecuencia:

$$\begin{aligned}\mathcal{E}(CM_{H(M)}) &= \sigma_e^2 + r\sigma_{H(M) \times Amb}^2 + ar\sigma_{H(M)}^2 = \sigma_e^2 + r\sigma_{H(M) \times Amb}^2 + ar(COV_{HC} - COV_{MH}) \\ &= \sigma_e^2 + r\sigma_{H(M) \times Amb}^2 + ar \left[\left(\frac{1}{2} \sigma_A^2 + \frac{1}{4} \sigma_D^2 \right) - \left(\frac{1}{4} \sigma_A^2 \right) \right] \\ &= \sigma_e^2 + r\sigma_{H(M) \times Amb}^2 + ar \left(\frac{1}{4} \sigma_A^2 + \frac{1}{4} \sigma_D^2 \right)\end{aligned}$$

Puede seguirse un razonamiento análogo para encontrar la expresión de las interacciones macho x ambiente y hembras (machos) x ambiente. Así:

$$\begin{aligned}\mathcal{E}(CM_{M^*Amb}) &= \sigma_e^2 + r\sigma_{H(M) \times Amb}^2 + rh\sigma_{M \times Amb}^2 = \sigma_e^2 + r\sigma_{H(M) \times Amb}^2 + rhCOV_{M \times Amb} \\ &= \sigma_e^2 + r\sigma_{H(M) \times Amb}^2 + rh \frac{1}{4} \sigma_{A \times Amb}^2\end{aligned}$$

y

$$\begin{aligned}\mathcal{E}(CM_{H(M)^*Amb}) &= \sigma_e^2 + r\sigma_{H(M) \times Amb}^2 = \sigma_e^2 + r(COV_{HC \times Amb} - COV_{MH \times Amb}) \\ &= \sigma_e^2 + r \left[\left(\frac{1}{2} \sigma_{A \times Amb}^2 + \frac{1}{4} \sigma_{D \times Amb}^2 \right) - \left(\frac{1}{4} \sigma_{A \times Amb}^2 \right) \right] \\ &= \sigma_e^2 + r \left(\frac{1}{4} \sigma_{A \times Amb}^2 + \frac{1}{4} \sigma_{D \times Amb}^2 \right)\end{aligned}$$

El sistema de ecuaciones constituido por las esperanzas de los cuadrados medios permite lograr estimaciones de las variancias de origen genético y de la heredabilidad.

Para estimar la variancia de "Machos" en ensayos en varios ambientes es necesario hacer una función lineal algo compleja. Puede estimarse a partir de

$$V_M = (CM_M - CMS)/arh, \text{ donde } CMS = CM_{M \times A} + CM_{H(M)} - CM_{H(M)^*A}$$

Hemos indicado más arriba que la variancia entre machos equivale a la covariancia de medios hermanos, por lo tanto:

$$V_M = \frac{1}{4} V_A = (CM_M - CMS)/arh$$

En consecuencia, la variancia aditiva estimada en la población de referencia será

$$V_A = 4V_M = 4(CM_M - CMS)/arh$$

La variancia de hembras dentro de machos puede estimarse como

$$V_{H(M)} = (CM_{H(M)} - CM_{H(M)^*Amb})/ar = \frac{1}{4} \sigma_A^2 + \frac{1}{4} \sigma_D^2$$

En consecuencia, la variancia genotípica total en la población de referencia, asumiendo ausencia de epistasis será:

$$V_G = 4 V_{H(M)} = V_A + V_D$$

Por lo tanto podemos lograr una estimación de la variancia de dominancia de la siguiente manera:

$$V_D = V_G - V_A = 4V_{H(M)} - 4V_M = 4(V_{H(M)} - V_M)$$

Siguiendo similar procedimiento podremos concluir que:

$$V_{A^*Amb} = 4V_{M^*Amb} = 4(CM_{M^*Amb} - CM_{H(M)^*Amb})/rh$$

$$V_{G^*Amb} = 4V_{H(M)^*Amb} = 4(CM_{H(M)^*Amb} - CM_{EE})/r$$

$$V_{D^*Amb} = 4V_{H(M)^*Amb} - 4V_{M^*Amb} = 4(V_{H(M)^*Amb} - V_{M^*Amb})$$

y

$$V_{EE} = CM_{EE}$$

Con las estimaciones logradas para la variancia del error, las de interacción genotipo x ambiente y las de origen genético es posible estimar las heredabilidades en sentido amplio y estricto sobre la base de plantas individuales en la población de referencia de la siguiente manera:

$$H = V_G / (V_G + V_{G^*Amb} + V_{EE}) = 4V_{H(M)} / (4V_{H(M)} + 4V_{H(M)xAmb} + V_{EE})$$

$$h^2 = V_A / (V_G + V_{G^*Amb} + V_{EE}) = 4V_M / (4V_{H(M)} + 4V_{H(M)xAmb} + V_{EE})$$

Las heredabilidades en sentido amplio y estricto para las medias de genotipos evaluados en a ambientes y r repeticiones dentro de ambientes serán (Hallauer y Miranda Filho, 1988)

$$H = V_G / (V_G + V_{G^*Amb}/a + V_{EE}/ar) = 4V_{H(M)} / (4V_{H(M)} + 4V_{H(M)xAmb}/a + V_{EE}/ar)$$

$$h^2 = V_A / (V_G + V_{G^*Amb}/a + V_{EE}/ar) = 4V_M / (4V_{H(M)} + 4V_{H(M)xAmb}/a + V_{EE}/ar)$$

8.5.3. Consideraciones sobre el diseño NC1

El esquema de cruzamientos de este diseño tiene la particularidad de adaptarse muy bien a los casos de especies alógamas, como por ejemplo maíz y girasol, entre los cultivos anuales. Ya que las plantas hembras están anidadas dentro de machos, en caso necesario es posible incrementar fácilmente el número de ellas para disponer de suficiente cantidad de semilla para evaluar en diferentes ambientes, a la vez que simultáneamente se mejora el muestreo de la población de referencia.

Desde el punto de vista de la estimación de las variancias de origen genético puede obtenerse una estimación de la variancia aditiva a partir de una ecuación que incluye solo dos cuadrados medios y por lo tanto la variabilidad del estimador puede ser relativamente baja. Por el contrario, la estimación de la variancia de dominancia es más compleja y con un nivel de precisión bastante menor al que se logra con un diseño NC2, como se verá más adelante.

Aplicando los conceptos indicados en la introducción, la variabilidad asociada a los estimadores de variancia aditiva y de variancia de dominancia será para el caso de evaluaciones en un ambiente, respectivamente:

$$V_{(VA)} = \{16 \times 2/r^2h^2\} \{[CM_M^2/(m + 1)] + [CM_{H(M)}^2/[m(h - 1) + 2]]\}$$

$$V_{(VD)} = \{16 \times 2/r^2h^2\} \{[CM_M^2/(m + 1)] + [(h + 1)^2 CM_{H(M)}^2/[m(h - 1) + 2] + [(h^2 CM_{EE}^2)/[(mh - 1)(r - 1) + 2]]\}$$

Nótese que el aumento en el número repeticiones (r), de machos (m) y de hembras dentro de machos (h) contribuye a mejorar la precisión de los estimadores de variancia aditiva y de dominancia, especialmente de la primera.

Cuando el experimento NC1 se realiza en varios ambientes, la forma de estimar las variancias aditiva y de dominancia son más complejas, de modo que la precisión de los estimadores será menor, aunque si las variancias de interacción aditiva x ambiente y dominante x ambiente fueran significativas, las estimaciones de la variancia aditiva y la de dominancia serían menos sesgadas.

8.5.4. Casos especiales del NC1 en función de los objetivos y de las características de la población de referencia

8.5.4.1. Extensión del esquema del NC1 al desarrollo y evaluación de líneas endocriadas

En un programa de mejoramiento, los individuos muestreados de la población, además de ser cruzados cada uno con diferentes individuos hembra, pueden ser autofecundados (si no existe autoincompatibilidad). La semilla $S_{0:1}$ de cada planta "macho" puede emplearse para iniciar el desarrollo de líneas endocriadas y la semilla de los cruzamientos empleada para evaluar la habilidad combinatoria general de cada planta S_0 utilizando la propia población como probadora.

8.5.4.2. Aplicación del NC1 en poblaciones de referencia derivadas del cruzamiento entre dos progenitores homocigotas

En este caso, las frecuencias génicas para los genes segregantes en la población de referencia serán $p = q = 0,5$. Como hemos visto anteriormente, asumiendo que no hay epistasis ni desequilibrio de ligamiento, en una población con $p=q=0,5$ es posible lograr una estimación del grado medio de dominancia (GMD) puesto que

$$\sigma_A^2 = (1/2)\sum a_i^2$$

$$\sigma_D^2 = (1/4)\sum d_i^2$$

Por lo tanto

$$V_M = \text{Cov}_{MH} = (1/4)V_A = (1/4)(1/2)\sum a_i^2$$

$$V_{H(M)} = \text{Cov}_{HC} - \text{Cov}_{MH} = (1/4)\sigma_A^2 + (1/4)\sigma_D^2 = (1/4)(1/2)\sum a_i^2 + (1/4)(1/4)\sum d_i^2$$

Y recordando que el GMD es:

$\text{GDM} = \sum d_i / \sum a_i$, podemos comprobar que

$$\begin{aligned} \text{GDM} &= [2(V_{H(M)} - V_M) / V_M]^{1/2} = [2(1/8\sum a_i^2 + 1/16\sum d_i^2 - 1/8\sum a_i^2) / 1/8\sum a_i^2]^{1/2} = \\ &= [(1/8)\sum d_i^2 / (1/8)\sum a_i^2]^{1/2} = \sum d_i / \sum a_i \end{aligned}$$

8.5.4.3. Aplicación del NC1 en poblaciones de referencia endocriadas

La variancia entre machos tendrá como esperanza de los cuadrados medios en evaluaciones en a ambientes a:

$$V_M = [(1 + F)/4]V_A$$

En consecuencia, la variancia aditiva estimada en la población de referencia será:

$$V_A = 4V_M / (1 + F)$$

En caso de tratarse de líneas con $F=1$,

$$V_M = (1/2)V_A$$

$$V_A = 2V_M$$

8.5.4.4. Consideraciones sobre epistasis en el diseño NC1

Hasta el momento asumimos epistasis ausente. En caso de que esta fuera distinta de cero y la población de referencia no fuera endocriada, las variancias de machos y de hembras dentro de macho serán:

$$V_M = \text{Cov}_{MH} = (1/4) V_A + (1/4)^2 V_{AA} + (1/4)^3 V_{AAA} + (1/4)^4 V_{AAAA} + \dots$$

Dentro de machos puede estimarse como

$$\begin{aligned} V_{H(M)} &= \text{Cov}_{HC} - \text{Cov}_{MH} = (1/2 V_A + 1/4 V_D + 1/4 V_{AA} + 1/8 V_{AD} + 1/16 V_{DD} \dots) - \\ &\quad - (1/4 V_A + (1/16) V_{AA} + (1/64) V_{AAA} + \dots) = \\ &= (1/4) V_A + (1/4) V_D + (3/16) V_{AA} + (1/8) V_{AD} + (1/16) V_{DD} \dots \end{aligned}$$

Por lo tanto:

$$\begin{aligned} 4V_M &= 4\text{Cov}_{MH} = V_A + (1/4) V_{AA} + (1/16) V_{AAA} + \dots \\ 4V_{H(M)} &= 4(\text{Cov}_{HC} - \text{Cov}_{MH}) = V_A + V_D + (3/4) V_{AA} + (1/2) V_{AD} + (1/4) V_{DD} \dots \\ 2(V_M + V_{H(M)}) &= 2[(1/4) V_A + (1/16) V_{AA} + \dots + (1/4) V_A + (1/4) V_D + (3/16) V_{AA} + \\ &\quad + (1/8) V_{AD} + (1/16) V_{DD} \dots] = \\ &= V_A + (1/2) V_D + (1/2) V_{AA} + (1/4) V_{AD} + (1/8) V_{DD} + \dots \end{aligned}$$

Estas tres expresiones poseen en común toda la variancia aditiva de la población de referencia. La comparación entre ellas permite tener una aproximación sobre la importancia relativa conjunta de los componentes de variancia de dominancia y de epistasia respecto de la variancia aditiva (Hallauer y Miranda Filho, 1988).

8.6. Diseño 2 de Carolina del Norte (NC2)

Este diseño propuesto por Comstock y Robinson (1948) requiere cruza-mientos entre individuos de la población de manera de generar progenies de medios hermanos y de hermanos completos. Como en todo diseño genético, tales individuos deben ser elegidos al azar de la población de referencia y en número suficiente para lograr una muestra representativa de la población para el carácter de interés. La particularidad del NC2 es que la estructura de tratamientos (genotipos) es de tipo cruzado (ver Anexo) porque cada uno de los individuos que se emplean como progenitores machos se cruzan todos ellos con las mismas hembras (ver esquema).

El esquema de estructura cruzada del diseño genético NC2 se presenta en el Cuadro 8.3.

M_1	M_2	M_3	...	M_m
H_1	H_1	H_1	...	H_1
H_2	H_2	H_2	...	H_2
H_3	H_3	H_3	...	H_3
H_4	H_4	H_4	...	H_4
H_5	H_5	H_5	...	H_5
...
H_f	H_f	H_f	H_f	H_f

Cuadro 8.3.

Esquema de estructura cruzada del diseño Carolina del Norte 2. M y H hacen referencia a los progenitores masculinos y femeninos.

Tal esquema da lugar a los tipos de progenies presentadas en el Cuadro 8.4.

M_1	M_2	M_3	...	M_m	
$HC_{1,1}$	$HC_{1,2}$	$HC_{1,3}$		$HC_{1,m}$	MH_1
$HC_{2,1}$	$HC_{2,2}$	$HC_{2,3}$		$HC_{2,m}$	MH_2
$HC_{3,1}$	$HC_{3,2}$	$HC_{3,3}$		$HC_{3,m}$	MH_3
...
$HC_{f,1}$	$HC_{f,2}$	$HC_{f,3}$		$HC_{f,m}$	MH_f
MH_1	MH_2	MH_3	...	MH_m	

Cuadro 8.4.

Tipos de familias de hermanos completos (HC) y de medios hermanos (MH) generados por el diseño Carolina del Norte 2.

Cada macho apareado con sus respectivas hembras da lugar a m familias de hermanos completos (HC). A su vez todas las progenies de hermanos completos que tienen un mismo macho en común dan lugar a m familias de medios hermanos paternos (MH) y todas las familias de hermanos completos que proceden de la misma hembra dan lugar a f familias de medios hermanos maternos (MH).

8.6.1. Análisis NC2 en un ambiente de evaluación

En el esquema que presenta la estructura de las familias vemos que tanto las progenies de cada macho como de cada hembra constituyen familias

de medios hermanos, de manera que las esperanzas de los cuadrados medios de "Machos" y de "Hembras" resultan ser:

$$\begin{aligned}\mathcal{E}(CM_M) &= \sigma_e^2 + r\sigma_{M^*H}^2 + rh\sigma_M^2 \\ &= \sigma_e^2 + r\sigma_{M^*H}^2 + rh^M COV_{MH} = \sigma_e^2 + r\sigma_{M^*H}^2 + rh \frac{1}{4} \sigma_A^2 \\ \mathcal{E}(CM_H) &= \sigma_e^2 + r\sigma_{M^*H}^2 + rm\sigma_H^2 = \sigma_e^2 + r\sigma_{M^*H}^2 + rm^H COV_{MH} \\ &= \sigma_e^2 + r\sigma_{M^*H}^2 + rm \frac{1}{4} \sigma_A^2\end{aligned}$$

Con el mismo razonamiento podemos entender la naturaleza de la interacción macho x hembra. Los cruzamientos entre machos y hembras constituyen familias de hermanos completos. Dentro de sí hay dos factores que disminuyen la variabilidad entre esas familias de hermanos completos, que son las circunstancias de que se trata de familias de medios hermanos por parte de cada macho y también por cada hembra común a los cruzamientos. En consecuencia:

$$\begin{aligned}\mathcal{E}(CM_{M^*H}) &= \sigma_e^2 + r\sigma_{M^*H}^2 = \sigma_e^2 + r(COV_{HC} - 2COV_{MH}) \\ &= \sigma_e^2 + r[(\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2) - 2(\frac{1}{4}\sigma_A^2)] \\ &= \sigma_e^2 + r(\frac{1}{4}\sigma_D^2)\end{aligned}$$

El esquema del análisis de variancia se presenta en la Tabla 8.9.

Fuentes de variación	Grados de libertad	Cuadrados medios (CM)	$\mathcal{E}(CM)$	F
Grupos	s-1			
Repeticiones (Grupos)	s(r-1)			
Machos (Grupos)	s(m-1)	CM_M	$\sigma_e^2 + r\sigma_{M^*H}^2 + rh\sigma_M^2$	CM_M / CM_{M^*H}
Hembras (Grupos)	s(h-1)	CM_H	$\sigma_e^2 + r\sigma_{M^*H}^2 + rm\sigma_H^2$	CM_H / CM_{M^*H}
[Machos * Hembras] (Grupos)	s(m-1)(h-1)	CM_{M^*H}	$\sigma_e^2 + r\sigma_{M^*H}^2$	CM_{M^*H} / CM_{EE}
Error experimental	s(r-1)(mh-1)	CM_{EE}	σ_e^2	
Total	srmh-1			

Tabla 8.9.

Esquema básico del análisis de variancia y esperanza de los cuadrados medios para un experimento NC2 con genotipos distribuidos aleatoriamente en s grupos con r repeticiones.

El sistema de ecuaciones constituido por las esperanzas de los cuadrados medios permite lograr estimaciones de las variancias de origen genético y de la heredabilidad.

$$V_H = (CM_H - CM_{M^*H})/rm$$

$$V_M = (CM_M - CM_{M^*H})/rh$$

Hemos indicado más arriba que las variancias entre machos y entre hembras equivalen a la covariancia de medios hermanos, por lo tanto:

$$V_M = 1/4 V_A$$

$$V_H = 1/4 V_A$$

El NC2 permite, entonces, obtener dos estimaciones independientes de la variancia aditiva de la población de referencia (V_A) y a los fines prácticos se utiliza el promedio de ambas estimaciones, a partir de "Machos" y a partir de "Hembras".

$${}^M V_A = 4V_M$$

$${}^H V_A = 4V_H$$

$$V_A = ({}^M V_A + {}^H V_A)/2 = 2(V_M + V_H)$$

La variancia de interacción machos x hembras puede estimarse a partir de:

$$V_{M^*H} = (CM_{M^*H} - CM_{EE})/r = 1/4 V_D$$

Por lo tanto, la variancia de dominancia (V_D) de la población de referencia será:

$$V_D = 4V_{M^*H}$$

$$V_D = 4(CM_{M^*H} - CM_{EE})/r$$

$$V_{EE} = CM_{EE}$$

Con las estimaciones logradas para la variancia del error, la de interacción genotipo x ambiente y las de origen genético, es posible estimar las heredabilidades en sentido amplio y estricto sobre la base de plantas individuales en la población de referencia de la siguiente manera:

$$H = (V_A + V_D) / (V_A + V_D + V_{EE})$$

$$= 2(V_M + V_H) + 4V_{M^*H} / [2(V_M + V_H) + 4V_{M^*H} + V_{EE}]$$

$$h^2 = V_A / (V_A + V_D + V_{EE})$$

$$= 2(V_M + V_H) / [2(V_M + V_H) + 4V_{M^*H} + V_{EE}]$$

Y las heredabilidades con base en la media de familias evaluadas en r repeticiones serán:

$$\begin{aligned}
 H &= (V_A + V_D) / [(V_A + V_D) + V_{EE}/r] \\
 &= 2(V_M + V_H) + 4V_{M^*H} / [2(V_M + V_H) + 4V_{M^*H} + V_{EE}/r] \\
 h^2 &= V_A / [V_A + V_D + V_{EE}/r] \\
 &= 2(V_M + V_H) / [2(V_M + V_H) + 4V_{M^*H} + V_{EE}/r]
 \end{aligned}$$

8.6.2. Análisis NC2 en varios ambientes de evaluación

El esquema del análisis de variancia para un diseño NC2 efectuado agrupando los genotipos aleatoriamente en s grupos o sets, en a ambientes y con r repeticiones dentro de grupos se presenta en Tabla 8.10.

Fuentes de variación	Grados de libertad	Cuadrados medios (CM)	ϵ (CM)	F
Ambientes	a-1			
Grupos	s-1			
Ambientes x Grupos	(a-1)(s-1)			
Repeticiones (Grupos)(Amb.)	as(r-1)			
Machos (Grupos)	s(m-1)	CM _M	$\sigma_e^2 + r\sigma_{M^*H^*A}^2 + rh\sigma_{M^*A}^2 + \frac{H}{ar\sigma_{M^*H}^2 + arh\sigma_M^2} (CM_M + CM_{M^*H^*A}) / (CM_{M^*A} + CM_{H^*A})$	
Hembras (Grupos)	s(h-1)	CM _H	$\sigma_e^2 + r\sigma_{M^*H^*A}^2 + rm\sigma_{H^*A}^2 + \frac{HH}{ar\sigma_{M^*H}^2 + arm\sigma_H^2} (CM_H + CM_{M^*H^*A}) / (CM_{M^*A} + CM_{H^*A})$	
[Machos x Hembras] (Grupos)	s(m-1)(h-1)	CM _{M^*H}	$\sigma_e^2 + r\sigma_{M^*H^*A}^2 + ar\sigma_{M^*H}^2$	CM _{M^*H} / CM _{M^*H^*A}
Machos (Grupos) x Amb.	s(m-1)(a-1)	CM _{M^*A}	$\sigma_e^2 + r\sigma_{M^*H^*A}^2 + rh\sigma_{M^*A}^2$	CM _{M^*A} / CM _{M^*H^*A}
Hembras (Grupos) x Amb.	s(h-1)(a-1)	CM _{H^*A}	$\sigma_e^2 + r\sigma_{M^*H^*A}^2 + rm\sigma_{H^*A}^2$	CM _{H^*A} / CM _{M^*H^*A}
[Machos*Hembras] (Grupos)*Amb	sa(m-1)(h-1)	CM _{M^*H^*A}	$\sigma_e^2 + r\sigma_{M^*H^*A}^2$	CM _{M^*H^*A} / CM _{EE}
Error experimental	sa(r-1)(mh-1)	CM _{EE}	σ_e^2	
Total	sarmh - 1			

^H y ^{HH} Los grados de libertad para la prueba son n_1 y n_2 , cada uno estimado de la siguiente manera (Hallauer y Miranda Filho, 1981):

$$n^1 = (CM_M + CM_{M^*H^*A})^2 / \{CM_M^2/[s(h+1)] + CM_{M^*H^*A}^2/[s(m-1)(h-1)(a-1)]\}$$

$$n^2 = (CM_{M^*A} + CM_{H^*A})^2 / \{CM_{M^*A}^2/[s(m-1)(a-1)] + CM_{H^*A}^2/[s(h-1)(a-1)]\}$$

Tabla 8.10.

Esquema básico del análisis de variancia y esperanza de los cuadrados medios para un experimento NC2 con genotipos distribuidos aleatoriamente en s grupos con r repeticiones y evaluados en a ambientes.

En el esquema que presenta la estructura de las familias, vemos que tanto las progenies de cada macho como de cada hembra constituyen familias de medios hermanos, de manera que la esperanza del cuadrado medio de "Machos" y de "Hembras" resultan ser:

$$\begin{aligned} \mathcal{E}(CM_M) &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 + rh\sigma_{M^*A}^2 + ar\sigma_{M^*H}^2 + arh\sigma_M^2 \\ &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 + rh\sigma_{M^*A}^2 + ar\sigma_{M^*H}^2 + arh^M COV_{MH} \\ &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 + rh\sigma_{M^*A}^2 + ar\sigma_{M^*H}^2 + arh \frac{1}{4} \sigma_A^2 \\ \mathcal{E}(CM_H) &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 + rm\sigma_{H^*A}^2 + ar\sigma_{M^*H}^2 + arm\sigma_H^2 \\ &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 + rm\sigma_{H^*A}^2 + ar\sigma_{M^*H}^2 + arm^H COV_{MH} \\ &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 + rm\sigma_{H^*A}^2 + ar\sigma_{M^*H}^2 + arm \frac{1}{4} \sigma_A^2 \end{aligned}$$

Con el mismo razonamiento podemos entender la naturaleza de la variabilidad machos x hembra. Los cruzamientos entre machos y hembras constituyen familias de hermanos completos. Dentro de sí hay dos factores que disminuyen la variabilidad entre esas familias de hermanos completos, que es la circunstancia de que se trata de familias de medios hermanos por parte de cada macho y cada hembra común a los cruzamientos.

En consecuencia:

$$\begin{aligned} \mathcal{E}(CM_{M^*H}) &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 + ar\sigma_{M^*H}^2 = \sigma_e^2 + r\sigma_{M^*H^*A}^2 + ar(COV_{HC} - 2COV_{MH}) \\ &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 + ar \left[\left(\frac{1}{2} \sigma_A^2 + \frac{1}{4} \sigma_D^2 \right) - 2 \left(\frac{1}{4} \sigma_A^2 \right) \right] \\ &= \sigma_e^2 + r\sigma_{H^*M^*A}^2 + ar \left(\frac{1}{4} \sigma_D^2 \right) \end{aligned}$$

Puede seguirse un razonamiento análogo para encontrar la expresión de las interacciones machos x ambiente, hembras x ambiente y machos x hembras x ambiente. Así:

$$\begin{aligned} \mathcal{E}(CM_{M^*A}) &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 + rh\sigma_{M^*A}^2 = \sigma_e^2 + r\sigma_{M^*H^*A}^2 + rhCOV_{M^*A} \\ &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 + rh \frac{1}{4} \sigma_{AxA}^2 \\ \mathcal{E}(CM_{H^*A}) &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 + rm\sigma_{H^*A}^2 = \sigma_e^2 + r\sigma_{M^*H^*A}^2 + rmCOV_{H^*A} \\ &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 + rm \frac{1}{4} \sigma_{AxA}^2 \\ \mathcal{E}(CM_{M^*H^*A}) &= \sigma_e^2 + r\sigma_{M^*H^*A}^2 = \sigma_e^2 + rCOV_{M^*H^*A} = \sigma_e^2 + r \frac{1}{4} \sigma_{AxA}^2 \end{aligned}$$

El sistema de ecuaciones constituido por las esperanzas de los cuadrados medios permite lograr estimaciones de las variancias de origen genético y de la heredabilidad.

Las variancias pueden estimarse a partir de:

$$V_{M^*H^*A} = (CM_{M^*H^*A} - CM_{EE})/r$$

$$V_{H^*A} = (CM_{H^*A} - CM_{M^*H^*A})/rm$$

$$V_{M^*A} = (CM_{M^*A} - CM_{M^*H^*A})/rh$$

$$V_H = (2CM_H - CM_{M^*H} - CM_{H^*A})/ram$$

$$V_M = (2CM_M - CM_{M^*H} - CM_{M^*A})/rah$$

Hemos indicado más arriba que la variancia entre machos y entre hembras equivale a la covariancia de medios hermanos, por lo tanto:

$$V_M = 1/4 V_A$$

$$V_H = 1/4 V_A$$

El NC2 permite obtener dos estimaciones independientes de la variancia aditiva de la población de referencia (V_A) y a los fines prácticos será el promedio de ambas estimaciones, a partir de "Machos" y a partir de "Hembras".

$${}^M V_A = 4V_M$$

$${}^H V_A = 4V_H$$

$$V_A = ({}^M V_A + {}^H V_A)/2 = 2(V_M + V_H)$$

La variancia de interacción machos x hembras puede estimarse a partir de:

$$V_{M^*H} = (CM_{M^*H} - CM_{M^*H^*A})/ar = 1/4 V_D$$

Por lo tanto, la variancia de dominancia (V_D) de la población de referencia será:

$$V_D = 4V_{M^*H}$$

$$V_D = 4(CM_{M^*H} - CM_{M^*H^*A})/ar$$

Siguiendo similares razonamientos podremos concluir que:

$$V_{M^*A} = 1/4 V_{A^*A}$$

$$V_{H^*A} = 1/4 V_{A^*A}$$

$$V_{M^*H^*A} = 1/4 V_{D^*A}$$

Entonces:

$${}^M V_{A^*A} = 4V_{M^*A}$$

$${}^H V_{A^*A} = 4V_{H^*A}$$

$$V_{A^*A} = ({}^M V_{A^*A} + {}^H V_{A^*A})/2 = 2(V_{M^*A} + V_{H^*A})$$

y

$$V_{D^*A} = 4V_{M^*H^*A}$$

$$V_{EE} = CM_{EE}$$

Con las estimaciones logradas para la variancia del error, las de interacción genotipo x ambiente y las de origen genético es posible estimar las heredabilidades en sentido amplio y estricto sobre la base de plantas individuales en la población de referencia de la siguiente manera:

$$\begin{aligned} H &= (V_A + V_D) / (V_A + V_D + V_{A^*A} + V_{D^*A} + V_{EE}) \\ &= [2(V_M + V_H) + 4V_{M^*H}] / [2(V_M + V_H) + 4V_{M^*H} + [2(V_{M^*A} + V_{H^*A}) + V_{EE}]] \\ h^2 &= V_A / (V_A + V_D + V_{A^*A} + V_{D^*A} + V_{EE}) \\ &= 2(V_M + V_H) / [2(V_M + V_H) + 4V_{M^*H} + [2(V_{M^*A} + V_{H^*A}) + V_{EE}]] \end{aligned}$$

Y las heredabilidades con base a la media de familias evaluadas en a ambientes y r repeticiones serán (Hallauer y Miranda Filho, 1988):

$$\begin{aligned} H &= (V_A + V_D) / [(V_A + V_D) + (V_{A^*A} + V_{D^*A})/a + V_{EE}/ra] \\ &= [2(V_M + V_H) + 4V_{M^*H}] / [2(V_M + V_H) + 4V_{M^*H} + 2(V_{M^*A} + V_{H^*A})/a + V_{EE}/ra] \\ h^2 &= V_A / [V_A + V_D + (V_{A^*A} + V_{D^*A})/a + V_{EE}/ra] \\ &= 2(V_M + V_H) / [2(V_M + V_H) + 4V_{M^*H} + 2(V_{M^*A} + V_{H^*A})/a + V_{EE}/ra] \end{aligned}$$

8.6.3. Consideraciones sobre el diseño NC2

El esquema de cruzamientos de este diseño no tiene la versatilidad que posee el diseño NC1. Aunque tampoco es de preferencia para especies autógamias por requerir la realización de cruzamientos, la característica del NC2 de requerir que los machos se apareen con las mismas hembras (diseño cruzado) hace necesario que la especie sea multiflora. Esa limitación no existiría si la población de referencia estuviera constituida por líneas cuyo F sea 1. Pero en tal caso el propósito general del NC2 no es estimar variancias genética y de dominancias, sino que se trataría de un diseño factorial cuyo objetivo básico es más bien evaluar las variancias de habilidad combinatoria general y de habilidad combinatoria específica. Volveremos sobre este tema al considerar los diseños dialélicos.

El diseño NC2 permite obtener dos estimaciones independientes de la variancia aditiva, en lugar de una sola como en el caso del NC1 y además permite una estimación directa de la variancia de dominancia (asumiendo que no existe variancia de interacción epistática).

En el caso de experimentos NC2 realizados en un único ambiente con r repeticiones, las variancias de los estimadores de variancia aditiva, a partir de machos y de hembras, serán, respectivamente:

$${}^M V_{(VA)} = \{16 \times 2/r^2 h^2\} \{[CM_M^2/(m+1)] + |CM_{M*H}^2/[(m-1)(h-1)+2]|\}$$

$${}^H V_{(VA)} = \{16 \times 2/r^2 m^2\} \{[CM_H^2/(h+1)] + |CM_{M*H}^2/[(m-1)(h-1)+2]|\}$$

Y la variancia de la estimación de la variancia de dominancia:

$$V_{(VD)} = \{16 \times 2/r^2\} \{[CM_{M*H}^2/[(m-1)(h-1)+2]] + |CM_{EE}^2/[(mh-1)(r-1)+2]|\}$$

El efecto de aumentar el número de machos y de hembras sobre la mejora de la precisión de los estimadores es aún mayor que en el diseño NC1 (obsérvese que estos factores aparecen solo en los denominadores de la fórmula de la variabilidad del estimador). En cuanto a las ventajas del NC2 respecto de las estimaciones sobre el NC1 nótese que se obtienen dos estimaciones independientes de la variancia aditiva y una estimación independiente de la variancia de dominancia.

Al igual que en el diseño NC1, cuando el experimento NC2 se realiza en varios ambientes, la forma de estimar las variancias aditivas y de dominancia son más complejas, de modo que la precisión de los estimadores será menor, aunque de resultar significativas las variancias de interacción aditiva \times ambiente y dominante \times ambiente resultarán menos sesgadas.

8.6.4. Casos especiales de aplicación de NC2 en función de los objetivos y de las características de la población de referencia

8.6.4.1. Extensión del esquema del NC2 al desarrollo y evaluación de líneas endocriadas

En un programa de mejoramiento, los individuos escogidos como machos de la población pueden, además de ser cruzados con una muestra de diferentes individuos hembra, ser autofecundados en caso de no existir autoincompatibilidad. De manera análoga al diseño NC1, la semilla $S_{0:1}$ de cada planta "macho" puede emplearse para iniciar el desarrollo de líneas endocriadas y la semilla de los cruzamientos empleada para evaluar la habilidad combinatoria general de cada planta S_0 utilizando un probador externo o la propia población como probadora.

8.6.4.2. Aplicación del NC2 en poblaciones de referencia derivadas del cruzamiento entre dos progenitores homocigotas

En este caso, las frecuencias génicas para los genes segregantes en la población de referencia serán $p=q=0,5$. Como hemos visto anteriormente, asumiendo que no hay epistasis ni desequilibrio de ligamiento, en una población con $p=q=0,5$, es posible lograr una estimación del grado promedio de dominancia puesto que:

$$\sigma_A^2 = (1/2)\sum a_i^2$$

$$\sigma_D^2 = (1/4)\sum d_i^2$$

Por lo tanto

$$V_M = V_H = \text{Cov}_{MH} = (1/4) V_A = (1/4)(1/2)\sum a_i^2 = (1/8)\sum a_i^2$$

$$\begin{aligned} V_{M^*H} &= \text{Cov}_{HC} - {}^M\text{Cov}_{MH} - {}^H\text{Cov}_{MH} = (1/2)V_A + (1/4)V_D - (1/4)V_A - (1/4)V_A \\ &= (1/4)V_D = (1/16)\sum d_i^2 \end{aligned}$$

Y recordando que el GMD es:

$$\text{GDM} = \sum d_i / \sum a_i$$

$$\text{GDM} = [2V_{M^*H}/V_M]^{1/2} = [(1/8)\sum d_i^2 / (1/8)\sum a_i^2]^{1/2} = \sum d_i / \sum a_i$$

8.6.4.3. Aplicación del NC2 en poblaciones de referencia endocriadas

Las variancias entre machos tendrán como esperanza de sus cuadrados medios a:

$$V_M = V_H = [(1+F)/4]V_A$$

$$V_{M^*H} = [(1+F)^2/2]V_D$$

En consecuencia, la variancia aditiva estimada en la población de referencia será:

$$V_A = 4V_M/(1+F) = 4V_H/(1+F)$$

En caso de tratarse de líneas con $F=1$,

$$V_M = V_H = (1/2)V_A$$

$$V_A = 2V_M = 2V_H \text{ o bien } V_A = V_M + V_H$$

8.6.4.4. Consideraciones sobre epistasis en el diseño NC2

Hasta el momento asumimos que no existe epistasis. En caso de que esta fuera distinta de cero y la población no fuera endocriada, las variancias de machos (o de hembras) y de machos x hembras serán:

$$V_M \text{ o } V_H = \text{Cov}_{MH} = \frac{1}{4} V_A + (1/16) V_{AA} + (1/64) V_{AAA} + \dots$$

La variancia de interacción macho x hembra puede estimarse como

$$\begin{aligned} V_{M*H} &= \text{Cov}_{HC} - {}^M\text{Cov}_{MH} - {}^H\text{Cov}_{MH} = [(\frac{1}{2}) V_A + (\frac{1}{4}) V_D + (1/4) V_{AA} + \\ &+ (1/8 V_{AD}) + 1/16 V_{DD} \dots] - [(\frac{1}{4}) V_A + (1/16) V_{AA} + \\ &+ (1/64) V_{AAA} + \dots] - [(\frac{1}{4}) V_A + (1/16) V_{AA} + (1/64) V_{AAA} + \dots] \\ &= (\frac{1}{4}) V_D + (1/8) V_{AA} + (1/8) V_{AD} + (1/16) V_{DD} \dots \end{aligned}$$

8.7. Diseño 3 de Carolina del Norte (NC3)

Este diseño fue concebido por Robinson *et al.* (1949) para indagar sobre la importancia relativa de la sobredominancia y si esta es real o producto del desequilibrio de ligamiento (pseudosobredominancia). La población de referencia es la F_2 del cruzamiento entre dos líneas homocigotas y si se asume que no existe desequilibrio de ligamiento, puede estimarse el grado promedio de dominancia. Hemos visto en el capítulo sobre variancias que cuando no hay equilibrio de ligamiento, las estimaciones de variancias genéticas y de dominancia se ven afectadas de manera diferente. La variancia de dominancia resulta sobreestimada cuando hay desequilibrio de ligamiento, y la variancia genética solo si el desequilibrio resulta en fase de acoplamiento. Si el desequilibrio es en fase de repulsión, la variancia genética estará subestimada respecto de la situación de equilibrio.

Individuos tomados al azar de la población F_2 se retrocruzan a ambas líneas progenitoras de la F_1 , utilizándolos como progenitores masculinos. Los genotipos para evaluar serán $2n + 2$ puesto que por cada individuo F_2 se generan dos progenies de medios hermanos que tienen en común a una u otra línea homocigota y además las dos líneas parentales. El esquema de cruzamientos se presenta en el Cuadro 8.5.

Machos	Hembras		
	Línea 1	Líneas 2	Medias
m_1	Y_{11}	Y_{12}	$Y_{1.}$
m_2	Y_{21}	Y_{22}	$Y_{2.}$
m_3	Y_{31}	Y_{32}	$Y_{3.}$
...
m_n	Y_{n1}	Y_{n2}	$Y_{n.}$
Medias	$Y_{.1}$	$Y_{.2}$	$Y_{..}$

Cuadro 8.5. Esquema de cruzamientos del diseño Carolina del Norte 3.

La tabla de ANdeVA para cada experimento de r repeticiones se presenta en la Tabla 8.11.

Fuentes de variación	Grados de libertad	Cuadrados medios (CM)	ϵ (CM)	F
Repeticiones	r-1			
Genotipos	2n-1			
Hembras (Líneas)	1	CM_H		
Machos	n-1	CM_M	$\sigma^2 + 2r\sigma_M^2 = \sigma^2 + \frac{1}{4}r\sum_i a_i^2$	$\frac{CM_M}{CM_{EE}}$; $\frac{CM_M}{CM_{MxH}}$
Machos x Hembras	n-1	CM_{MxH}	$\sigma^2 + r\sigma_{MxH}^2 = \sigma^2 + \frac{1}{4}r\sum_i d_i^2$	CM_{MxH} / CM_{EE}
Error experimental	(r-1)(2n-1)	CM_{EE}	σ^2	
Total	2rn - 1			

Tabla 8.11. Esquema del análisis de variancia para un experimento NC3 evaluado en un único ambiente con r repeticiones.

La variancia entre machos es la covariancia de medios hermanos que tienen en común al n-ésimo macho. La variancia entre hembras es la covariancia entre medios hermanos que tienen como progenitor común a la línea 1 o a la línea 2.

La variancia de interacción Macho x Hembra es la covariancia entre hermanos completos, menos la suma de las covariancias de medios hermanos a

través de machos y de líneas (estas últimas diferentes entre sí, a diferencia de lo que ocurre en el NC2).

Las variancias entre machos y de machos x hembras se estiman a partir de las esperanzas de los cuadrados medios de la siguiente manera:

$$V_M = \text{Cov}_{MH} = 1/4 V_A = (CM_M - CM_{EE})/2r = 1/2 \sum_i p_i q_i a_i^2$$

En la población de referencia (F_2), las frecuencias génicas son $p_i=q_i=0,5$, consecuentemente V_M

$$V_M = 1/2 \sum_i (0,5)(0,5) a_i^2 = 1/8 \sum_i a_i^2, \quad [8.1]$$

y

$$V_{MxH} = (CM_{MxH} - CM_{EE})/r = \sum_i p_i q_i d_i^2 = V_D = 1/4 \sum_i d_i^2, \quad [8.2]$$

Por definición, (ver Capítulo 3) una estimación del grado medio de dominancia (GMD) puede lograrse como:

$$\text{GMD} = \sum_i d_i^2 / \sum_i a_i^2$$

Teniendo en consideración [8.1] y [8.2],

$$\text{GMD} = (V_{MxH}/2V_M)^{1/2}$$

Las pruebas F de significancia de los cuadrados medios de Machos y de Machos x Hembras se calculan como el cociente entre los cuadrados medios respectivos divididos por el cuadrado medio del error experimental (CM_{EE}). Si la prueba de F resulta significativa para Machos x Hembra, ello indica que existe dominancia puesto que el término $\sum_i d_i^2$ sería diferente de cero. La prueba F CM_M/CM_{MH} sirve para determinar si la dominancia es completa o está en el nivel de parcial o sobredominante. Si la relación CM_M/CM_{MH} no es significativamente distinta de 1, ello se interpreta como que la dominancia es completa. Si la relación es significativamente menor a 1 implica sobredominancia y si es mayor a 1 implica dominancia parcial (o falta de dominancia si CM_{MxH}/CM_{EE} no hubiese resultado significativo).

En caso de existir, el desequilibrio de ligamiento será máximo en la F_2 . La reproducción de la población F_2 mediante varios ciclos de apareamiento aleatorio (Syn-n) irá disminuyendo tal desequilibrio. Para determinar si la sobredominancia en la F_2 para el carácter de interés era real, o simplemente producto de la afectación de las variancias de dominancia y de aditividad por el desequilibrio de ligamiento (pseudosobredominancia), el diseño NC3 debe repetirse sobre la nueva población de referencia: F_2 Syn-n.

Como requerimiento general de los diseños genéticos, también el NC3 precisa un muestreo satisfactorio, aleatorio y con número de individuos suficiente, de la población de referencia. Por ello es necesario distribuir aleatoriamente a las progenies en sets o grupos. La tabla de ANDeVA en el caso de experimentos NC3 replicados en varios ambientes se presenta en la Tabla 8.12.

Fuentes de variación	Grados de libertad	Cuadrados medios (CM)	ε (CM)	F
Ambientes	a-1			
Grupos (Amb.)	a(g-1)			
Repeticiones (Grupos)(Amb)	ag(r-1)			
Hembras (Grupos)	g	CM _H	$\sigma^2 + r n \sigma_{AxH}^2 + a r n \sigma_H^2$	CM _H / CM _{AxH}
Hembras (Grupos) x Amb.	g(a-1)	CM _{AxH}	$\sigma^2 + r n \sigma_{AxH}^2$	CM _{AxH} / CM _{EE}
Machos (Grupos)	g(n-1)	CM _M	$\sigma^2 + 2r \sigma_{AxM}^2 + 2a r \sigma_M^2$	CM _M / CM _{AxM}
Machos (Grupos) x Amb.	g(n-1)(a-1)	CM _{AxM}	$\sigma^2 + 2r \sigma_{AxM}^2$	CM _{AxM} / CM _{EE}
Machos(G) x Hembras (G)	g(n-1)	CM _{MxH}	$\sigma^2 + r \sigma_{AxMxH}^2 + a r \sigma_{MxH}^2$	CM _{MxH} / CM _{AxMxH}
Machos(G) x Hembras (G) x Amb.	g(n-1)(a-1)	CM _{AxMxH}	$\sigma^2 + r \sigma_{AxMxH}^2$	CM _{AxMxH} / CM _{EE}
Error experimental	ag(r-1)(2n-1)	CM_{EE}	σ²	
Total	2agr-1			

Tabla 8.12. Esquema del análisis de variancia para un experimento NC3 evaluado en con r repeticiones y a ambientes.

A partir del NC3 se pueden lograr estimaciones de la heredabilidad en el sentido estricto en la población F₂ sobre la base de medias de parcelas:

$$h^2 = 4\sigma_M^2 / [4\sigma_M^2 + 4\sigma_{AxM}^2/a + \sigma_{MxH}^2 + \sigma_{AxMxH}^2/a + \sigma^2/ar],$$

sobre la base de plantas individuales:

$$h^2 = 4\sigma_M^2 / [4\sigma_M^2 + 4\sigma_{AxM}^2 + \sigma_{MxH}^2 + \sigma_{AxMxH}^2 + \sigma^2],$$

o sobre la base de medias de familias de medios hermanos:

$$h^2 = \sigma_M^2 / [\sigma_M^2 + \sigma_{AxM}^2/a + \sigma_{MxH}^2 + \sigma_{AxMxH}^2/a + \sigma^2/ar],$$

8.8. Diseños dialélicos

Este tipo de diseños genéticos son ampliamente utilizados en los programas de mejoramiento, particularmente en los de especies alógamas bajo los supuestos del modelo de efectos fijos y preferentemente con la finalidad de estimar efectos de habilidad combinatoria general y específica de líneas endocriadas. Bajo el modelo aleatorio, puede emplearse para estimar los componentes de variancias de habilidad combinatoria general y de habilidad combinatoria específica de poblaciones de cría. La limitación principal de los diseños dialélicos para estimar variancias genotípicas es la dificultad para incluir un número de individuos que permita muestrear razonablemente bien la variabilidad genotípica en una población de referencia. En el esquema general del diseño dialélico, todos los progenitores deben cruzarse entre sí en todas las combinaciones posibles (Tabla 8.13) y el número de combinaciones posibles aumenta en forma exponencial con el aumento del número de aquellos.

Número de progenitores	Cruzamientos rectos,recíprocos y progenitores (M1)	Cruzamientos directos y progenitores (M2)	Cruzamientos directos y recíprocos (M3)	Cruzamientos directos (M4)
(n)	(n ²)	[n(n-1)/2 +n]	[n(n-1)]	[n(n-1)/2]
3	9	6	6	3
4	16	10	12	6
5	25	15	20	10
6	36	21	30	15
7	49	28	42	21
8	64	36	56	28
9	81	45	72	36
10	100	55	90	45
15	225	120	210	105
20	400	200	380	180
25	625	325	600	300
50	2500	1275	2450	1225
100	10000	5050	9900	4950
200	40000	20100	39800	19900

Tabla 8.13.

Número de genotipos para evaluar en función del número de genotipos progenitores, considerando todos los cruzamientos simples, o todos los cruzamientos simples y sus recíprocos.

Un muestreo adecuado de una población de referencia hace que el número de individuos requerido sea muy grande. Tanto en un modelo fijo como en uno aleatorio, un número grande de progenitores exige realizar y evaluar un número elevado de cruzamientos que puede resultar impracticable. Aunque ello impone restricciones al análisis, una alternativa es el uso de cruzamientos dialélicos parciales, donde no es preciso disponer de todas las combinaciones de cruzamientos posibles entre un grupo de progenitores (Miranda Filho y Geraldi, 1984). La organización del germoplasma en grupos permite disminuir el número de cruzamientos. En tal caso el diseño adopta un esquema factorial en el que un grupo de n líneas se cruzan con n' líneas pertenecientes a otro grupo. En tal caso, el número de híbridos para evaluar se reduce considerablemente (Tabla 8.14).

Progenitores $n = n' + n''$	Número de progenitores		Número de genotipos para evaluar	
	Genotipos progenitores del grupo A (n')	Genotipos progenitores del grupo B (n'')	Cruzamientos directos $[n(n-1)/2]$	Cruzamientos directos entre grupos $n'n''$
4	2	2	6	4
10	5	5	45	25
10	6	4	45	24
10	7	3	45	21
20	10	10	190	100
20	11	9	190	99
20	12	8	190	96
20	13	7	190	91
20	14	6	190	84
50	25	25	1225	625
50	30	20	1225	600
50	35	15	1225	525
50	40	10	1225	400
200	100	100	19900	10000
200	120	80	19900	9600
200	140	60	19900	8400
200	170	30	19900	5100

Tabla 8.14. Número de cruzamientos para evaluar en función del número de progenitores agrupados en dos clases de n' y n'' integrantes, considerando todos los cruzamientos simples.

Una excelente revisión general de los diferentes esquemas y análisis de experimentos dialélicos, sus alcances, limitaciones y equivalencias entre sí puede encontrarse en Baker (1978).

8.8.1. Diseño dialélico de Griffing (1956)

Los diseños dialélicos de Griffing son una familia de experimentos con cuatro variantes o métodos (1 a 4) según cuales sean los genotipos para evaluar. Con las limitaciones señaladas más arriba, pueden emplearse para estimar efectos (modelo de efectos fijos) o variancias (modelo de efectos aleatorios) de habilidad combinatoria general (HCG) y habilidad combinatoria específica (HCE). Sobre el tema puede revisarse la sección correspondiente del capítulo 4. Los diferentes esquemas de genotipos para evaluar correspondientes a los métodos de Griffing se presentan en la Tabla 8.15, y los modelos estadísticos correspondientes a cada uno en la Tabla 8.16, donde:

g_i y g_j es la habilidad combinatoria general de los progenitores i y j ,

s_{ij} es la habilidad combinatoria específica del cruzamiento entre los progenitores i x j

r_{ij} es el efecto recíproco en el cruzamiento i x j ,
donde $r_{ij} = m_i + m_j + n_{ij}$, siendo m_i el efecto materno del progenitor i -ésimo, m_j el del progenitor j -ésimo y n_{ij} el efecto no materno del cruzamiento i x j

e_{ijk} es el error asociado a la k -ésima observación del cruzamiento i x j

Método 1. Dialélico completo. Incluye los progenitores y los cruzamientos en ambos sentidos

Progenitores	1	2	3	4	5
1	1x1	1x2	1x3	1x4	1x5
2	2x1	2x2	2x3	2x4	2x5
3	3x1	3x2	3x3	3x4	3x5
4	4x1	4x2	4x3	4x4	4x5
5	5x1	5x2	5x3	5x4	5x5

Método 2. Incluye los progenitores y los cruzamientos en uno solo de los sentidos

Progenitores	1	2	3	4	5
1	1x1	1x2	1x3	1x4	1x5
2		2x2	2x3	2x4	2x5
3			3x3	3x4	3x5
4				4x4	4x5
5					5x5

Método 3. Excluye los progenitores, e incluye los cruzamientos en ambos sentidos

Progenitores	1	2	3	4	5
1		1x2	1x3	1x4	1x5
2	2x1		2x3	2x4	2x5
3	3x1	3x2		3x4	3x5
4	4x1	4x2	4x3		4x5
5	5x1	5x2	5x3	5x4	

Método 4. Incluye únicamente los cruzamientos en uno solo de los sentidos

Progenitores	1	2	3	4	5
1		1x2	1x3	1x4	1x5
2			2x3	2x4	2x5
3				3x4	3x5
4					4x5
5					

Tabla 8.15. Esquemas de cruzamientos dialélicos y genotipos para evaluar en los cuatro métodos propuestos por Griffing (1956).



Método	Modelo	Restricciones
1	$Y_{ijk} = u + g_i + g_j + s_{ij} + r_{ij} + e_{ijk}$	$\sum_i g_i = 0; \sum_{ij} s_{ij} = 0; \sum_{ij} r_{ij} = 0$
2	$Y_{ijk} = u + g_i + g_j + s_{ij} + e_{ijk}$	$\sum_i g_i = 0; \sum_{ij} s_{ij} = 0;$
3	$Y_{ijk} = u + g_i + g_j + s_{ij} + r_{ij} + e_{ijk}$	$\sum_i g_i = 0; \sum_{ij} s_{ij} = 0; \sum_i r_{ij} = 0$
4	$Y_{ijk} = u + g_i + g_j + s_{ij} + e_{ijk}$	$\sum_i g_i = 0; \sum_{ij} s_{ij} = 0;$

Tabla 8.16.

Modelos genéticos del diseño dialélico de Griffing (1956) para los cuatro métodos propuestos y restricciones de sus correspondientes modelos.

Wricke y Weber (1986) desaconsejan la utilización de los métodos 1 y 2 de Griffing para estimar variancias (modelo aleatorio) debido además a que en estos métodos deben evaluarse los progenitores endocriados. Asumiendo que se trata de un diseño dialélico de Griffing donde se prueban los $n(n-1)/2$ cruzamientos posibles entre n líneas endocriadas (método 4), las medias observadas (Y_{ij}) pueden disponerse en una tabla de doble entrada, con una columna adicional que contiene las medias marginales, es decir, las medias de todos los cruzamientos que tienen en común al n -ésimo progenitor y en la última fila la media general (Tabla 8.17).

Progenitores	2	3	4	5	Totales y medias marginales
1	Y_{1x2}	Y_{1x3}	Y_{1x4}	Y_{1x5}	$Y_{1x.} ; (Y_{1x2} + Y_{1x3} + Y_{1x4} + Y_{1x5}) / (n-1)$
2		Y_{2x3}	Y_{2x4}	Y_{2x5}	$Y_{2x.} ; (Y_{1x2} + Y_{2x3} + Y_{2x4} + Y_{2x5}) / (n-1)$
3			Y_{3x4}	Y_{3x5}	$Y_{3x.} ; (Y_{1x3} + Y_{2x3} + Y_{3x4} + Y_{3x5}) / (n-1)$
4				Y_{4x5}	$Y_{4x.} ; (Y_{1x4} + Y_{2x4} + Y_{3x4} + Y_{4x5}) / (n-1)$
5					$Y_{5x.} ; (Y_{1x5} + Y_{2x5} + Y_{3x5} + Y_{4x5}) / (n-1)$
Total y media general	$Y_{..} ; (Y_{1x2} + Y_{1x3} + Y_{1x4} + Y_{1x5} + Y_{2x3} + Y_{2x4} + Y_{2x5} + Y_{3x4} + Y_{3x5} + Y_{4x5}) / [n(n-1)/2]$				

Tabla 8.17.

Procedimiento de cómputo de los totales y medias marginales, total y media general en un experimento dialélico de Griffing (1956), método 4.

Los desvíos entre las medias marginales para cada progenitor y la media general son, por definición, las habilidades combinatorias generales (HCG) de cada uno de los n progenitores de los cruzamientos. Las habilidades combinatorias específicas (HCE) son los desvíos entre las medias observadas para cada cruzamiento ixj y la suma de las HCG de los progenitores i y j . La variación de HCE es la variación entre celdas dentro de márgenes. Cada margen corresponde a los cruzamientos que tienen un padre común.

A manera de ejemplo, los efectos de HCG de los progenitores 3 y 5 serán:

$$HCG_3 = g_3 = Y_{3x.} - Y_{..}$$

$$HCG_5 = g_5 = Y_{5x.} - Y_{..}$$

y el efecto de HCE entre los progenitores 3 y 5 será:

$$HCE_{3x5} = s_{3x5} = Y_{3x5} - (g_3 + g_5) - Y_{..}$$

Cuando se realiza un diseño dialélico de Griffing y se prueban los $n(n-1)/2$ cruzamientos posibles entre n líneas endocriadas y los progenitores (método 2) las medias observadas (Y_{ixj}) pueden disponerse en una tabla de doble entrada, con una columna adicional que contiene las medias marginales, es decir, las medias de todo los cruzamientos que tienen en común a la n -ésima (Tabla 8.18).

Progenitores	1	2	3	4	5	Totales y medias marginales
1	Y_{1x1}	Y_{1x2}	Y_{1x3}	Y_{1x4}	Y_{1x5}	$Y_{1x.}; (Y_{1x1} + Y_{1x2} + Y_{1x3} + Y_{1x4} + Y_{1x5})/n$
2		Y_{2x2}	Y_{2x3}	Y_{2x4}	Y_{2x5}	$Y_{2x.}; (Y_{1x2} + Y_{2x2} + Y_{2x3} + Y_{2x4} + Y_{2x5})/n$
3			Y_{3x3}	Y_{3x4}	Y_{3x5}	$Y_{3x.}; (Y_{1x3} + Y_{2x3} + Y_{3x3} + Y_{3x4} + Y_{3x5})/n$
4				Y_{4x4}	Y_{4x5}	$Y_{4x.}; (Y_{1x4} + Y_{2x4} + Y_{3x4} + Y_{4x4} + Y_{4x5})/n$
5					Y_{5x5}	$Y_{5x.}; (Y_{1x5} + Y_{2x5} + Y_{3x5} + Y_{4x5} + Y_{5x5})/n$
Total y media general	$Y_{..}; (Y_{1x1} + Y_{2x2} + Y_{3x3} + Y_{4x4} + Y_{5x5} + Y_{1x2} + Y_{1x3} + Y_{1x4} + Y_{1x5} + Y_{2x3} + Y_{2x4} + Y_{2x5} + Y_{3x4} + Y_{3x5} + Y_{4x5}) / \{[n(n-1)/2] + n\}$					

Tabla 8.18. Procedimiento de cómputo de los totales y medias marginales y total y media general en un experimento dialélico de Griffing (1956), método 2.

En el caso de los métodos donde se disponen los cruzamientos en ambos sentidos (directos y recíprocos), los efectos recíprocos se estiman como:

$$r_{ij} = (Y_{ixj} - Y_{jxi})/2,$$

y los efectos maternos como:

$$m_i = (Y_{.i} - Y_{.i})/2n$$

Para experimentos realizados en un ambiente con r repeticiones siguiendo el método Griffing 4 la tabla de ANdeVa toma la configuración que se presenta en la Tabla 8.19.

Fuentes de variación	Grados de libertad	Cuadrados medios (CM)	Esperanza de los cuadrados medios (CM)		F
			Modelo I	Modelo II	
Repeticiones	$r-1$				
Cruzamientos	$[n(n-1)/2]-1$	CM_C	$\sigma^2 + rK_C^2$	$\sigma^2 + r\sigma_C^2$	CM_C / CM_{EE}
H.C.G.	$n-1$	CM_{HCG}	$\sigma^2 + r[(n-2)/(n-1)]K_{HCG}^2$	$\sigma^2 + r\sigma_{HCE}^2 + r(n-2)\sigma_{HCG}^2$	CM_{HCG} / CM_{EE} CM_{HCG} / CM_{HCE}
H.C.E.	$n(n-3)/2$	CM_{HCE}	$\sigma^2 + r[2/n(n-3)]K_{HCE}^2$	$\sigma^2 + r\sigma_{HCE}^2$	CM_{HCE} / CM_{EE}
Error exper.	$(r-1)[n(n-1)/2-1]$	CM_{EE}	σ^2	σ^2	
Total	$[rn(n-1)/2]-1$				

Tabla 8.19.

Esquema del análisis de variancia para un experimento dialélico con r repeticiones en un único ambiente siguiendo el modelo de Griffing, método 4.

Cuando la fuente de variación “cruzamientos” es significativa, resulta útil subdividir la suma de cuadrados en forma ortogonal en dos sumas de cuadrados, la debida a la habilidad combinatoria general y la debida a la habilidad combinatoria específica. De acuerdo al modelo que se emplee, las pruebas de significancia F difieren en el denominador para utilizar como error. En el modelo de efectos fijos, la significancia de los efectos de habilidad combinatoria general y específica se prueban con el cuadrado medio del error experimental, mientras que bajo el modelo aleatorio, la significancia de la variancia de HCE se prueba con el cuadrado medio del error experimental y la de la variancia de HCG con el cuadrado medio de la HCE, a menos que esta hubiera resultado no significativa. En este último caso puede emplearse directamente como error el cuadrado medio del error experimental.

Las variancias de HCG y de HCE pueden expresarse en términos de componentes de variancia entre parientes. La variancia de HCG es la variancia entre cruzamientos que tienen un progenitor en común, es decir, la variancia entre medios hermanos:

$$\sigma_{\text{HCG}}^2 = \text{Cov}_{\text{G(MH)}} = [(1 + F)/4] \sigma_A^2 \text{ (en ausencia de epistasis y para niveles arbitrarios de endocría de los progenitores).}$$

Si los progenitores del dialélico son líneas completamente homocigotas, $F=1$ y por lo tanto:

$$\sigma_{\text{HCG}}^2 = \text{Cov}_{\text{G(MH)}} = 1/2 \sigma_A^2 \quad [8.3]$$

La variancia de HCE es la variancia entre cruzamientos dentro de líneas. La variancia de cruzamientos es la variancia entre hermanos completos (combinaciones línea x línea), pero está "reducida" por el hecho de que cada uno de los progenitores ha sido utilizado como progenitor común de $n-1$ cruzamientos. Por lo tanto y asumiendo que no hay epistasis:

$$\begin{aligned} \sigma_{\text{HCE}}^2 &= \text{Cov}_{\text{G(HC)}} - 2\text{Cov}_{\text{G(MH)}} = [(1 + F)/2] \sigma_A^2 + [(1 + F)/2]^2 \sigma_D^2 - 2[(1 + F)/4] \sigma_A^2 \\ &= [(1 + F)/2]^2 \sigma_D^2 \end{aligned}$$

En el caso de que los progenitores sean líneas completamente homocigotas ($F=1$):

$$\sigma_{\text{HCE}}^2 = \text{Cov}_{\text{G(HC)}} - 2\text{Cov}_{\text{G(MH)}} = \sigma_A^2 + \sigma_D^2 - 2(1/2 \sigma_A^2) = \sigma_D^2 \quad [8.4]$$

Si en cambio los progenitores no fueran endocriados ($F=0$):

$$\sigma_{\text{HCG}}^2 = \text{Cov}_{\text{G(MH)}} = 1/4 \sigma_A^2 \quad [8.5]$$

$$\sigma_{\text{HCE}}^2 = \text{Cov}_{\text{G(HC)}} - 2\text{Cov}_{\text{G(MH)}} = 1/4 \sigma_D^2 \quad [8.6]$$

Las estimaciones de las variancias aditiva y de dominancia de la población de referencia pueden obtenerse a partir de considerar las esperanzas de los cuadrados medios. La Tabla 8.20 presenta el esquema de ANdeVA para el caso de un diseño método 4 de Griffing, modelo II.



Fuentes de variación	Grados de libertad	Cuadrados medios (CM)	Esperanza de cuadrados medios
			Modelo aleatorio (II)
Repeticiones	r-1		
Cruzamientos	[n(n-1)/2]-1	CM _C	$\sigma^2 + \sigma_c^2$
H.C.G.	n - 1	CM _{HCG}	$\sigma^2 + r[\text{Cov}_{G(HC)} - 2 \text{Cov}_{G(MH)}] + r(n-2)\text{Cov}_{G(MH)}$
H.C.E.	n(n-3)/2	CM _{HCE}	$\sigma^2 + r[\text{Cov}_{G(HC)} - 2 \text{Cov}_{G(MH)}]$
Error exper.	(r-1)[n(n-1)/2-1]	CM _{EE}	σ^2
Total	[rn(n-1)/2]-1		

Tabla 8.20.

Esquema del ANDeVA para un cruzamiento dialélico siguiendo a Griffing (1956), método 4, considerando un modelo aleatorio.

Si F=1, considerando [8.3] y [8.4], respectivamente:

$$V_A = V_{HCG} = 2(CM_{HCG} - CM_{HCE})/r(n-2)$$

$$V_D = V_{HCE} = (CM_{HCE} - CM_{EE})/r$$

Si F=0, considerando [8.5] y [8.6], respectivamente:

$$V_A = 4V_{HCG} = 4(CM_{HCG} - CM_{HCE})/r(n-2)$$

$$V_D = 4V_{HCE} = 4(CM_{HCE} - CM_{EE})/r$$

8.8.1.1. Método 1

En el caso del método 1, los arreglos y medias marginales serán los que se presentan en la Tabla 8.21.

Progenitores	1	2	3	4	5	Totales y medias marginales
1	Y ₁₁	Y ₁₂	Y ₁₃	Y ₁₄	Y ₁₅	Y _{1.} ; (Y ₁₁ + Y ₁₂ + Y ₁₃ + Y ₁₄ + Y ₁₅)/n
2	Y ₂₁	Y ₂₂	Y ₂₃	Y ₂₄	Y ₂₅	Y _{2.} ; (Y ₁₂ + Y ₂₂ + Y ₂₃ + Y ₂₄ + Y ₂₅)/n
3	Y ₃₁	Y ₃₂	Y ₃₃	Y ₃₄	Y ₃₅	Y _{3.} ; (Y ₁₃ + Y ₂₃ + Y ₃₃ + Y ₃₄ + Y ₃₅)/n
4	Y ₄₁	Y ₄₂	Y ₄₃	Y ₄₄	Y ₄₅	Y _{4.} ; (Y ₁₄ + Y ₂₄ + Y ₃₄ + Y ₄₄ + Y ₄₅)/n
5	Y ₅₁	Y ₅₂	Y ₅₃	Y ₅₄	Y ₅₅	Y _{5.} ; (Y ₁₅ + Y ₂₅ + Y ₃₅ + Y ₄₅ + Y ₅₅)/n
Totales marginales, total y media general	Y _{.1}	Y _{.2}	Y _{.3}	Y _{.4}	Y _{.5}	Y _{..} ; (Y ₁₁ + Y ₁₂ + Y ₁₃ + Y ₁₄ + Y ₁₅ + Y ₂₁ + Y ₂₂ + Y ₂₃ + Y ₂₄ + Y ₂₅ + Y ₃₁ + Y ₃₂ + Y ₃₃ + Y ₃₄ + Y ₃₅ + Y ₄₁ + Y ₄₂ + Y ₄₃ + Y ₄₄ + Y ₄₅ + Y ₅₁ + Y ₅₂ + Y ₅₃ + Y ₅₄ + Y ₅₅)/n ²

Tabla 8.21.

Totales y medias marginales, total y media general correspondientes al método 1 de Griffing (1956), ejemplificado para 5 progenitores.

El ANdeVA se presenta en Tabla 8.22.

Fuentes de variación	Grados de libertad	Cuadrados medios (CM)	Esperanza de cuadrados medios	
			Modelo fijo	Modelo aleatorio
Entre progenitores	n-1	CM _p	$\sigma^2 + rK_p^2$	$\sigma^2 + r\sigma_p^2$
Entre cruzamientos	n(n-1)-1	CM _c	$\sigma^2 + rK_c^2$	$\sigma^2 + r\sigma_c^2$
H.C.G.	n - 1	CM _{HCG}	$\sigma^2 + r[2n/(n-1)]K_{HCG}^2$	$\sigma^2 + r2(n-1)\sigma_{HCE}^2 + 2rn\sigma_{HCG}^2$
H.C.E.	n(n-1)/2	CM _{HCE}	$\sigma^2 + r [2/[n(n-1)]K_{HCE}^2$	$\sigma^2 + [r2(n^2-n+1)/n^2]\sigma_{HCE}^2$
Recíprocos	n(n-1)/2	CM _R	$\sigma^2 + [r4/n(n-1)]K_R^2$	$\sigma^2 + r\sigma_R^2$
Error experim.	(r-1)(n ² -1)		σ^2	σ^2

Tabla 8.22.

Claves del ANdeVA correspondientes a un análisis dialélico aplicando el método 1, según un modelo de efectos fijos y de efectos aleatorios. Adaptado de Cruz y Regazzi, 1994.

Las estimaciones de la media general y de los efectos de habilidad combinatoria general, habilidad combinatoria específica y recíprocos, respectivamente, se calculan como (Cruz y Regazzi, 1994):

$$u = Y_{..} / n^2$$

$$g_i = [(Y_{i.} + Y_{.i}) / 2n] - Y_{..} / n^2$$

$$s_{ij} = (Y_{ij} + Y_{ji}) / 2 - (Y_{i.} + Y_{.i} + Y_{.j} + Y_{j.}) / 2n + Y_{..} / n^2$$

$$r_{ij} = (Y_{ij} - Y_{ji}) / 2$$

y las sumas de cuadrados de HCG, HCE y efectos recíprocos se calculan como:

$$SC_{HCG} = [\sum_i (Y_{i.} + Y_{.i})^2] / 2n - Y_{..}^2 / n^2$$

$$SC_{HCE} = \sum_i \sum_j Y_{ij} (Y_{ij} + Y_{ji}) / 2 - \sum (Y_{i.} + Y_{.i})^2 / 2n + Y_{..}^2 / n^2$$

$$SC_R = (1/2) \sum \sum_{i < j} (Y_{ij} - Y_{ji})^2$$

8.8.1.2. Método 2

En el caso del método 2, los arreglos y medias marginales serán los que se presentan en la Tabla 8.23

Progenitores	1	2	3	4	5	Totales marginales
1	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	$Y_{1.}$
2		Y_{22}	Y_{23}	Y_{24}	Y_{25}	$Y_{2.}$
3			Y_{33}	Y_{34}	Y_{35}	$Y_{3.}$
4				Y_{44}	Y_{45}	$Y_{4.}$
5					Y_{55}	$Y_{5.}$
Totales marginales	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$	$Y_{.4}$	$Y_{.5}$	Total general $Y_{..}$

Tabla 8.23.

Totales marginales y general correspondientes al método 2 de Griffing (1956), ejemplificado para 5 progenitores.

Las estimaciones de la media general y de los efectos de habilidad combinatoria general y habilidad combinatoria específica, respectivamente, se calculan como (Cruz y Regazzi, 1994):

$$u = [1/n(n + 1)]Y_{..}$$

$$g_i = [1/(n + 2)][(Y_{i.} + Y_{.i}) - (2/n)Y_{..}]$$

$$s_{ij} = Y_{ij} - 2(Y_{i.} + Y_{.j})/(n+2) + 2Y_{..}/[(n + 1)(n + 2)]$$

Fuentes de variación	Grados de libertad	Cuadrados medios (CM)	ϵ (CM)	
			Mod. fijo (I)	Mod. aleatorio (II)
Repeticiones	$r-1$			
Genotipos	$[n(n+1)]/2-1$	CM_G	$\sigma^2 + rK_G^2$	$\sigma^2 + r\sigma_G^2$
Progenitores vs. Cruzamientos	1	CM_{PvsC}	$\sigma^2 + rK_{PvsC}^2$	$\sigma^2 + r\sigma_{PvsC}^2$
Entre progenitores	$n-1$	CM_P	$\sigma^2 + rK_P^2$	$\sigma^2 + r\sigma_P^2$
Entre cruzamientos	$[n(n-1)]/2 - 1$	CM_C	$\sigma^2 + rK_C^2$	$\sigma^2 + r\sigma_C^2$
H.C.G.	$n - 1$	CM_L	$\sigma^2 + r[(n+2)/(n-1)]K_{HCG}^2$	$\sigma^2 + r\sigma_{HCE}^2 + r(n-2)\sigma_{HCG}^2$
H.C.E.	$n(n-3)/2$	$CM_{C/L}$	$\sigma^2 + r[2n(n-3)]K_{HCE}^2$	$\sigma^2 + r\sigma_{HCE}^2$
Error experimental	$(r-1)[n(n+1)/2-1]$	CM_{EE}	σ^2	σ^2
Total	$r[n(n+1)/2]-1$			

Tabla 8.24.

Claves del ANDeVA correspondientes a un análisis dialélico aplicando el método 2, según un modelo de efectos fijos y de efectos aleatorios.

El ANDeVA se presenta en Tabla 8.24. La misma contiene la partición ortogonal que proponen Hallauer y Miranda Filho (1988) que presenta el contraste entre progenitores y cruzamientos, de manera de tener otra prueba sobre la significancia de la acción génica no aditiva.

En el análisis sin la variante mencionada, las sumas de cuadrados de HCG y HCE se calculan de acuerdo a Cruz y Regazzi (1994) como:

$$SC_{HCG} = [\sum_i (Y_i + Y_{ii})^2 - (4Y_{..}^2/n)] / (n+2)$$

$$SC_{HCE} = \sum_{i < j} \sum Y_{ij}^2 - [\sum_i (Y_i + Y_{ii})^2] / (n+2) + [2Y_{..}^2 / (n^2 + 3n + 2)]$$

8.8.1.3. Método 3

En el método 3 los arreglos serán como se indica en Tabla 8.25:

Progenitores	1	2	3	4	5	Totales marginales
1		Y_{12}	Y_{13}	Y_{14}	Y_{15}	$Y_{.1}$
2	Y_{21}		Y_{23}	Y_{24}	Y_{25}	$Y_{.2}$
3	Y_{31}	Y_{32}		Y_{34}	Y_{35}	$Y_{.3}$
4	Y_{41}	Y_{42}	Y_{43}		Y_{45}	$Y_{.4}$
5	Y_{51}	Y_{52}	Y_{53}	Y_{54}		$Y_{.5}$
Totales marginales	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$	$Y_{.4}$	$Y_{.5}$	Total general $Y_{..}$

Tabla 8.25. Totales marginales y general correspondientes al método 3 de Griffing (1956), ejemplificado para 5 progenitores

Las sumas de cuadrados de HCG y HCE se calculan (Cruz y Regazzi, 1994) como:

$$SC_{HCG} = [1/2(n - 2)] \sum_i (Y_i + Y_{.i})^2 - [2/(n(n - 2))] Y_{..}^2$$

$$SC_{HCE} = 1/2 \sum_{i < j} \sum (Y_{ij} + Y_{ji})^2 - [1/2(n - 2)] \sum_i (Y_i + Y_{.i})^2 + [1/(n - 1)(n - 2)] Y_{..}^2$$

$$SC_R = 1/2 \sum_{i < j} \sum (Y_{ij} - Y_{ji})^2$$

Las estimaciones de la media general, de los efectos de habilidad combinatoria general, de la habilidad combinatoria específica y de los recíprocos, respectivamente, se calculan como:

$$u = [1/n(n + 1)] Y_{..}$$

$$g_i = [1/[2n(n-2)]] [n(Y_i + Y_{.i}) - 2Y_{..}]$$

$$s_{ij} = \frac{1}{2} (Y_{ij} + Y_{ji}) - \frac{[(Y_{i.} + Y_{.j} + Y_{i.} + Y_{.j}) / [2(n-2)]] + Y_{..} / (n^2 - 3n + 2)}$$

$$r_{ij} = (Y_{ij} - Y_{ji}) / 2$$

En el caso de modificar el modelo para discriminar los efectos maternos, estos se estiman como:

$$m_i = (Y_{i.} + Y_{.i}) / 2n$$

El ANdeVA se presenta en las Tablas 8.26 y 8.27

Fuentes de variación	Grados de libertad	Cuadrados medios (CM)	E(CM)	
			Modelo fijo (I)	Modelo aleatorio (II)
Repeticiones	r-1			
Entre cruzam.	n(n-1)-1	CM _G	$\sigma^2 + rK_G^2$	$\sigma^2 + r\sigma_G^2$
H.C.G.	n - 1	CM _{HCG}	$\sigma^2 + r[2(n-2)/(n-1)]K_{HCG}^2$	$\sigma^2 + 2r\sigma_{HCE}^2 + r2(n-2)\sigma_{HCG}^2$
H.C.E.	n(n-3)/2	CM _{HCE}	$\sigma^2 + r[4/(n-3)]K_{HCE}^2$	$\sigma^2 + r2\sigma_{HCE}^2$
Efectos recípr.	n(n-1)/2	CM _R	$\sigma^2 + r[4/n(n-1)]K_R^2$	$\sigma^2 + r\sigma_R^2$
Error experim.	(r-1)[(n-1)-1]	CM _{EE}	σ^2	σ^2
Total	rn(n-1)-1			

Tabla 8.26.

Claves del ANdeVA correspondientes a un análisis dialélico aplicando el método 3, según un modelo de efectos fijos y de efectos aleatorios.

Fuentes de variación	Grados de libertad	Cuadrados medios (CM)	E(CM)
			Modelo aleatorio (II)
Repeticiones	r-1		
Genotipos	n(n-1)-1	CM _G	$\sigma^2 + r\sigma_G^2$
H.C.G.	n - 1	CM _{HCG}	$\sigma^2 + 2r\sigma_{HCE}^2 + 2r(n-2)\sigma_{HCG}^2$
H.C.E.	n(n-3)/2	CM _{HCE}	$\sigma^2 + 2r\sigma_{HCE}^2$
Efectos maternos	n-1	CM _M	$\sigma^2 + rn r\sigma_M^2$
Efectos recíprocos	(n-1)(n-2)/2	CM _R	$\sigma^2 + r\sigma_R^2$
Error experimental	(r-1)[(n-1)]-1	CM _{EE}	σ^2
Total	rn(n-1)-1		

Tabla 8.27.

Claves del ANdeVA correspondientes a un análisis dialélico aplicando el método 3, considerando efectos maternos.

8.8.1.4. Método 4

En el método 4 los arreglos serán según se indica en Tabla 8.28

Progenitores	1	2	3	4	5	Totales marginales
1		Y_{12}	Y_{13}	Y_{14}	Y_{15}	$Y_{1.}$
2			Y_{23}	Y_{24}	Y_{25}	$Y_{2.}$
3				Y_{34}	Y_{35}	$Y_{3.}$
4					Y_{45}	$Y_{4.}$
5						$Y_{5.}$
Media y total general	$(Y_{12} + Y_{13} + Y_{23} + Y_{14} + Y_{24} + Y_{34} + Y_{15} + Y_{25} + Y_{35} + Y_{45})/10$					$Y_{..}$

Tabla 8.28. Totales marginales, media y total general correspondientes al método 4 de Griffing (1956), ejemplificado para 5 progenitores.

Las estimaciones de la media general y de los efectos de habilidad combinatoria general, habilidad combinatoria específica, respectivamente, se calculan como:

$$u = [2/n(n - 1)]Y_{..}$$

$$g_i = [1/n(n - 2)][nY_{i.} - 2Y_{..}]$$

$$s_{ij} = Y_{ij} - [1/(n - 2)][Y_{i.} + Y_{.j}] + \{2/[(n - 1)(n - 2)]\}Y_{..}$$

Las Tablas 8.29 y 8.30 presentan el esquema del análisis de variancia de un experimento dialélico método 4 realizado en uno y en varios ambientes, respectivamente.

Fuentes de variación	Grados de libertad (gl)	Cuadrados medios (CM)	E(CM)	
			Modelo fijo (I)	Modelo aleatorio (II)
Repeticiones	$r-1$			
Cruzamientos	$n(n-1)/2 - 1$	CM_C	$\sigma^2 + rK_C^2$	$\sigma^2 + r\sigma_C^2$
H.C.G.	$n - 1$	CM_{HCG}	$\sigma^2 + r[(n-2)/(n-1)]K_{HCG}^2$	$\sigma^2 + r\sigma_{HCE}^2 + r(n-2)\sigma_{HCG}^2$
H.C.E.	$n(n-3)/2$	CM_{HCE}	$\sigma^2 + r\{2/[n(n-3)]\}K_{HCE}^2$	$\sigma^2 + r\sigma_{HCE}^2$
Error experim.	$(r-1)[n(n-1)/2-1]$	CM_{EE}	σ^2	σ^2
Total	$rn(n-1)/2-1$			

Tabla 8.29. Claves del ANDeVA correspondientes a un análisis dialélico aplicando el método 4 en un ambiente.



Fuentes de variación	Grados de libertad	Cuadrado medio (CM)	E(CM)	
			Modelo fijo (I)*	Modelo aleatorio (II)
Ambientes	a-1			
Repeticiones/ Amb.	a(r-1)			
Cruzamientos	n(n-1)/2 -1	CM _C	$\sigma^2 + arK_C^2$	$\sigma^2 + r\sigma_{CXA}^2 + ra\sigma_C^2$
H.C.G.	n - 1	CM _{HCG}	$\sigma^2 + r[a(n-2)/n-1]K_{HCG}^2$	$\sigma^2 + r\sigma_{HCEXA}^2 + ra\sigma_{HCE}^2 + r(n-2)\sigma_{HCGXA}^2 + ra(n-2)\sigma_{HCG}^2$
H.C.E.	n(n-3)/2	CM _{HCE}	$\sigma^2 + r\{2a/[n(n-3)]\}K_{HCE}^2$	$\sigma^2 + r\sigma_{HCEXA}^2 + ra\sigma_{HCE}^2$
Cruz. x Amb.	(a-1)[n(n-1)/2]-1	CM _{CxA}	$\sigma^2 + r\sigma_{CxA}^2$	$\sigma^2 + r\sigma_{CxA}^2$
H.C.G. x Amb.	(a-1)(n-1)	CM _{HCGxA}	$\sigma^2 + r[(n-2)/(n-1)]\sum K_{HCGxA}^2$	$\sigma^2 + r\sigma_{HCEXA}^2 + r(n-2)\sigma_{HCGxA}^2$
H.C.E. x Amb.	(a-1)[n(n-3)/2]	CM _{HCEXA}	$\sigma^2 + r[2/n(n-3)]\sum K_{HCEXA}^2$	$\sigma^2 + r\sigma_{HCEXA}^2$
Error exp.	a(r-1)[n(n-1)/2]-1	CM _{EE}	σ^2	σ^2
Total	[arn(n-1)/2]-1			

*Se considera que las repeticiones y los ambientes son variables aleatorias.

Tabla 8.30.

Claves del ANDeVA correspondientes a un análisis dialélico aplicando el método 4 combinado a través de ambientes.

Seguindo a Cruz y Regazzi (1994), las sumas de cuadrados de HCG y HCE se calculan como:

$$SC_{HCG} = [1/(n - 2)] \sum_i Y_i^2 - [4/(n(n - 2))]Y_{..}^2$$

$$SC_{HCE} = \sum_{i<j} \sum Y_{ij}^2 - [1/(n - 2)]\sum_i Y_i^2 + [2/(n - 1)(n - 2)]Y_{..}^2$$

8.8.2. Variancia de los efectos

En la Tabla 8.31 se presentan las variancias de los efectos y de la diferencia entre efectos correspondientes al análisis dialélico de Griffing, métodos 1 a 4 (Cruz y Regazzi, 1994).

Efecto o diferencia entre efectos	Método 1	Método 2	Método 3	Método 4
u	$(1/n^2)\sigma^2$	$[2/n(n+2)]\sigma^2$	$[1/n(n-1)]\sigma^2$	$[2/n(n-1)]\sigma^2$
g_i	$[(n-1)/2n^2]\sigma^2$	$[(n-1)/n(n+2)]\sigma^2$	$\{(n-1)/[2n(n-2)]\}\sigma^2$	$[(n-1)/n(n-2)]\sigma^2$
$g_i - g_j$	$(1/n)\sigma^2$	$[2/(n+2)]\sigma^2$	$[1/(n-2)]\sigma^2$	$[2/(n-2)]\sigma^2$
s_{ij}	$[(n^2-2n+2)/2n^2]\sigma^2$	$[(n^2+n+2)/[(n+1)(n+2)]\sigma^2$	$\{(n-3)/[2(n-1)]\}\sigma^2$	$[(n-3)/(n-1)]\sigma^2$
$s_{ij}-s_{ik}$	$[(n-1)/n]\sigma^2$	$[2(n+1)/(n+2)]\sigma^2$	$[(n-3)/(n-2)]\sigma^2$	$[2(n-3)/(n-2)]\sigma^2$
$s_{ij}-s_{km}$	$[(n-2)/n]\sigma^2$	$[2n/(n+2)]\sigma^2$	$[(n-4)/(n-2)]\sigma^2$	$[2(n-4)/(n-2)]\sigma^2$
r_{ij}	$(1/2)\sigma^2$		$(1/2)\sigma^2$	
$r_{ij} - r_{km}$	$(1)\sigma^2$		$(1)\sigma^2$	

Tabla 8.31.

Variancia de la media general, efectos y diferencia entre efectos (g , habilidad combinatoria general; s , habilidad combinatoria específica; r , recíprocos) correspondientes a los cuatro métodos dialécticos de Griffing (1956) con n líneas parentales. Adaptado de Cruz y Regazzi (1994).

8.8.4. Importancia relativa de las habilidades combinatorias general y específica

Para expresar la importancia relativa de las variancias y efectos de habilidad combinatoria general y específica en la determinación del comportamiento de los híbridos entre líneas endocriadas Baker (1978) propuso utilizar las siguientes expresiones:

$$2\sigma_{HCG}^2 / (2\sigma_{HCG}^2 + \sigma_{HCE}^2),$$

o bien

$$2K_{HCG}^2 / (2K_{HCG}^2 + K_{HCE}^2)$$

Cuanto más cercana a 1 resulte la relación, mayor será la importancia de la habilidad combinatoria general como determinante del comportamiento de los cruzamientos híbridos.

8.9. Diseños dialélicos de Gardner y Eberhart

8.9.1. Análisis I

Gardner y Eberhart (1966), Gardner y Lonquist (1966), y Gardner (1967) propusieron varias modalidades de análisis. El análisis I para el cruzamiento dialélico requiere la evaluación de cinco clases genotípicas:

- Las n poblaciones progenitoras o variedades.
- Las n poblaciones progenitoras o variedades luego de una generación de autofecundación.
- Los $n(n - 1)/2$ cruzamientos interpoblacionales en todas las combinaciones posibles, sin distinguir recíprocos.
- Los $n(n - 1)/2$ cruzamientos interpoblacionales seguidos de una generación de autofecundación.
- Los $n(n - 1)/2$ cruzamientos interpoblacionales seguidos de una generación de apareamiento aleatorio.

Ello permite estimar los parámetros ya vistos (ver Capítulo 4):

$$\mu, A_i, D_i, A_j, D_j, P_h, h_i, h_j \text{ y } s_{ij}$$

Pueden plantearse cinco modelos genético-estadísticos que difieren en las constantes consideradas y que dan lugar a las sumas de cuadrados que se indican en la Tabla 8.32 (adaptado de Gardner y Eberhart, 1966).

Modelo	Constantes	Sumas de cuadrados
1	μ, A_i	SC_1
2	μ, A_i, D_i	SC_2
3	μ, A_i, D_i, P_h	SC_3
4	μ, A_i, D_i, P_h, h_i	SC_4
5	$\mu, A_i, D_i, P_h, h_i, s_{ij}$	SC_5

Tabla 8.32. Parámetros genéticos a estimar en el análisis I del modelo dialélico de Gardner y Eberhart (1966).

Las sumas de cuadrados de las fuentes de variación en el ANdeVA pueden calcularse a partir de diferencias entre las sumas de cuadrados de los diferentes modelos indicados más arriba (Tabla 8.33).

Fuente de variación	Grados de libertad	Sumas de cuadrados
Poblaciones	$[n(3n+1)/2] - 1$	SC_p
Aditividad	$n-1$	$SC_A = SC_1 - FC$
Dominancia	n	$SC_D = SC_2 - SC_1$
Heterosis	$n(n-1)/2$	$SC_H = SC_5 - SC_2$
Heterosis Promedio	1	$SC_{HP} = SC_3 - SC_2$
Heterosis Varietal	$n-1$	$SC_{HV} = SC_4 - SC_3$
Heterosis Específica	$n(n-3)/2$	$SC_{HS} = SC_5 - SC_4$
Residual	$n(n-1)$	$SC_R = SC_p - (SC_A + SC_D + SC_H)$

Tabla 8.33. Esquema del ANdeVA para un experimento dialélico de Gardner y Eberhart (1966), análisis I.

El análisis I permite obtener información sobre la contribución de los genotipos homocigotas y los heterocigotas a la media poblacional, así como sobre la heterosis y sus componentes. El residual del modelo acumula efectos sobre las medias de las variedades debidos eventualmente a la epistasis y al ligamiento.

8.9.2. Análisis II

Estos diseños tienen por objeto analizar los cruzamientos entre poblaciones y requieren la evaluación de las n poblaciones *per se* y de todos sus cruzamientos interpoblacionales (se asume que no hay efectos recíprocos significativos, por lo cual se evaluarán $n(n-1)/2$ híbridos intervarietales). Por ejemplo, para un dialélico de cinco ($n=5$) poblaciones, el esquema de los cruzamientos y arreglos toman la configuración que se presenta en la Tabla 8.34.

Pob.	P ₁	P ₂	P ₃	P ₄	P ₅	Y _i (o Y _i)	Y _{hi}	MY _{hi}
P ₁	Y ₁₁	Y ₁₂	Y ₁₃	Y ₁₄	Y ₁₅	Y _{1.} = Y ₁₁ + Y ₁₂ + Y ₁₃ + Y ₁₄ + Y ₁₅	Y _{1.} - Y ₁₁	(Y _{1.} - Y ₁₁)/(n-1)
P ₂	Y ₂₁	Y ₂₂	Y ₂₃	Y ₂₄	Y ₂₅	Y _{2.} = Y ₂₁ + Y ₂₂ + Y ₂₃ + Y ₂₄ + Y ₂₅	Y _{2.} - Y ₂₂	(Y _{2.} - Y ₂₂)/(n-1)
P ₃	Y ₃₁	Y ₃₂	Y ₃₃	Y ₃₄	Y ₃₅	Y _{3.} = Y ₃₁ + Y ₃₂ + Y ₃₃ + Y ₃₄ + Y ₃₅	Y _{3.} - Y ₃₃	(Y _{3.} - Y ₃₃)/(n-1)
P ₄	Y ₄₁	Y ₄₂	Y ₄₃	Y ₄₄	Y ₄₅	Y _{4.} = Y ₄₁ + Y ₄₂ + Y ₄₃ + Y ₄₄ + Y ₄₅	Y _{4.} - Y ₄₄	(Y _{4.} - Y ₄₄)/(n-1)
P ₅	Y ₅₁	Y ₅₂	Y ₅₃	Y ₅₄	Y ₅₅	Y _{5.} = Y ₅₁ + Y ₅₂ + Y ₅₃ + Y ₅₄ + Y ₅₅	Y _{5.} - Y ₅₅	(Y _{5.} - Y ₅₅)/(n-1)

$$Y_v = Y_{11} + Y_{22} + Y_{33} + Y_{44} + Y_{55}$$

$$Y_H = Y_{12} + Y_{13} + Y_{14} + Y_{15} + Y_{23} + Y_{24} + Y_{25} + Y_{34} + Y_{35} + Y_{45}$$

$$Y_{..} = Y_v + Y_H$$

Tabla 8.34. Cruzamientos y arreglos para un experimento dialéctico de Gardner y Eberhart (1966) siguiendo al análisis II.

El diseño de Gardner y Eberhart guarda similitud con el diseño dialéctico de Griffing, método 2, bajo modelo de efectos fijos (1), pero es más informativo por cuanto subdivide adicionalmente los efectos de habilidad combinatoria específica en tres componentes.

El modelo estadístico que describe la media de los genotipos (variedades o poblaciones) a evaluar es:

$$Y_{ij} = \mu_v + \frac{1}{2} (v_i + v_j) + \Theta h_{ij}$$

donde:

μ_v es la media de las poblaciones.

v_i o v_j son los efectos de las variedades progenitoras de cada híbrido o cruzamiento intervarietal.

h_{ij} son los efectos de heterosis del híbrido intervarietal entre los progenitores i y j.

Θ es una constante igual a 1 cuando $i \neq j$ (híbrido) o igual a 0 cuando $i = j$ (población progenitora).

De tal modo para las poblaciones progenitoras evaluadas *per se*, la expresión resulta:

$$Y_{ii} = \mu_v + \frac{1}{2} (v_i + v_i) + 0h_{ij} = \mu_v + v_i$$

El modelo requiere estimar n efectos varietales y $n(n-1)/2$ efectos de heterosis.

El efecto de heterosis (h_{ij}) puede descomponerse de la siguiente forma:

$$h_{ij} = {}^P h + h_i + h_j + s_{ij},$$

donde:

${}^P h$ es la heterosis promedio;

h_i es la heterosis varietal de la i -ésima población progenitora;

h_j es la heterosis varietal de la j -ésima población progenitora;

s_{ij} es el efecto de heterosis específica del cruzamiento entre la i -ésima y j -ésima población progenitora.

De manera similar al análisis I, podemos plantear la serie de modelos estadísticos que se presentan en la Tabla 8.35.

Modelo	Constantes	Suma de cuadrados
1	$Y_{ij} = \mu_v + \frac{1}{2} (v_i + v_j)$	SC ₁
2	$Y_{ij} = \mu_v + \frac{1}{2} (v_i + v_j) + \theta {}^P h$	SC ₂
3	$Y_{ij} = \mu_v + \frac{1}{2} (v_i + v_j) + \theta {}^P h + \theta (h_i + h_j)$	SC ₃
4	$Y_{ij} = \mu_v + \frac{1}{2} (v_i + v_j) + \theta {}^P h + \theta (h_i + h_j) + \theta s_{ij}$	SC ₄

Tabla 8.35. Parámetros genéticos para estimar en el análisis II del modelo dialélico de Gardner y Eberhart (1966).

Para el cálculo de los efectos debemos presentar la estructura de los datos, a partir de la tabla anterior y con las consideraciones respecto a que no existen efectos recíprocos, las constantes del modelo se estiman como se indica en la Tabla 8.36 y el esquema del análisis de variancia en la Tabla 8.37.

	Constante	Estimador
Media de todos los progenitores	μ_V	Y_V/n
Efecto varietal	V_i	$Y_{ii} - Y_V$
Efecto de heterosis	$h_{ij} = P_h + h_i + h_j + s_{ij}$	$Y_{ij} - (Y_{ii} + Y_{jj})/2$
<i>Heterosis promedio</i>	P_h	$Y_H - Y_V$
<i>Heterosis varietal</i>	h_i	$[(n-1)/(n-2)](M Y_{hi} - Y_H) - \frac{1}{2} V_i$
<i>Heterosis específica</i>	s_{ij}	$Y_{ij} + [n/(n-2)]Y_H - [(n-1)/(n-2)](M Y_{hi} + M Y_{hj})$

Tabla 8.36. Parámetros y cálculo de sus estimadores para un experimento dialélico de Gardner y Eberhart (1966) según análisis II.

Fuente de variación	Grados de libertad	Sumas de cuadrados	Cuadrados medios	F
Genotipos	$[n(n+1)/2] - 1$	SC_G	CM_G	
Progenitores	$n-1$	$SC_P = SC_1 - FC$	CM_P	CM_P/CM_{EE}
Heterosis	$n(n-1)/2$	$SC_H = SC_4 - SC_1$	CM_H	CM_H/CM_{EE}
<i>Heterosis promedio</i>	1	$SC_{HP} = SC_2 - SC_1$	CM_{HP}	CM_{HP}/CM_{EE}
<i>Heterosis varietal</i>	$n-1$	$SC_{HV} = SC_3 - SC_2$	CM_{HV}	CM_{HV}/CM_{EE}
<i>Heterosis específica</i>	$n(n-3)/2$	$SC_{HS} = SC_4 - SC_3$	CM_{HS}	CM_{HS}/CM_{EE}
Error experimental	$(r-1)\{[n(n+1)/2]-1\}$	SC_{EE}	CM_{EE}	

Tabla 8.37. Esquema del análisis II de un experimento dialélico según Gardner y Eberhart (1966).

A los fines prácticos, en el caso de que ciertos componentes de la heterosis resulten no significativos, se recomienda eliminarlos del modelo, ganando mayor precisión en las estimaciones de los restantes efectos. A diferencia del análisis I, no resulta posible discriminar los efectos de aditividad y de dominancia.

8.9.3. Análisis III

En este análisis el conjunto de genotipos evaluados incluye a las poblaciones progenitoras o variedades y a los cruzamientos. A diferencia de los anteriores análisis, la media de los cruzamientos intervarietales se expresa según el siguiente modelo:

$$Y_{ij} = \mu_c + g_i + g_j + s_{ij}$$

Esto lo asemeja al dialélico según Griffing, método 4 y modelo de efectos fijos. Obsérvese que este modelo no incorpora términos que correspondan a los efectos varietales, aunque se incluyan en el diseño experimental a las variedades progenitoras.

El modelo para las poblaciones progenitoras es

$$Y_{ii} = \mu_v + v_i$$

El ANdeVA según el análisis III sigue el esquema de la Tabla 8.38.

Fuentes de variación	Grados de libertad	Sumas de cuadrados
Variedades	n-1	SC _v
Variedades vs. Cruzamientos	1	SC _{vvs.c}
Cruzamientos	$[n(n-1)/2] - 1$	SC _c
<i>Habilidad combinatoria general</i>	n-1	SC _{HCG}
<i>Habilidad combinatoria específica</i>	n(n-3)/2	SC _{HCE}
Error experimental		

Tabla 8.38. Esquema del análisis de variancia para un experimento dialélico, análisis III de Gardner y Eberhart (1966).

8.9.4. Relación entre los análisis II y III

Existe correspondencia entre algunas de las fuentes de variación:

La SC_v en el análisis II es igual a la SC_{vvs.c} en el análisis III, mientras que la suma (SC_v + SC_{HV}) en el análisis II equivale a la suma (SC_v + SC_{HCG}) en el análisis III.

8.10. Procedimientos para estimar los componentes de variancias poblacionales

El análisis de los componentes de las variancias fenotípica y genotípica puede llevarse a cabo mediante dos estrategias diferentes. Una es el análisis de variancia (ANdeVA) y el otro es la utilización de modelos de máxima verosimilitud restringida (REML). La preferencia por uno u otro de estos esquemas depende de las características de los datos disponibles. En el caso de datos desbalanceados, las estimaciones de componentes de variancia utilizando ANdeVA tienden a resultar sesgadas. Debido a que el desbalance de los datos es muy frecuente, los modelos REML son los más utilizados por cuanto proveen de mejores estimaciones de los componentes de variancia y de mejores predictores insesgados (BLUPs) del mérito genético de los genotipos con fines de selección (Piepho *et al.*, 2008). Estos temas serán tratados en los próximos capítulos.

En los programas de mejoramiento, aun cuando los datos sean balanceados, es frecuente que las poblaciones estén en desequilibrio de ligamiento. La misma selección genera desequilibrio genético y consecuentemente las estimaciones obtenibles pueden estar sesgadas (ver Capítulo 5).

Los métodos de estimación de componentes de variancia a partir de la resolución simultánea de sistemas de ecuaciones construidos sobre las esperanzas de los cuadrados medios no resultan ser correctos cuando los datos presentan desbalance, es decir, que existen diferencias en la cantidad de información disponible entre los niveles de una variable clasificatoria, por efectos de parcelas perdidas, repeticiones perdidas, o aun pérdidas de ambientes completos.

Cuando ello ocurre debe emplearse otro tipo de modelos. El modelo de máxima verosimilitud (*maximum likelihood*, ML) fue concebido y desarrollado por Fisher (1921) y Hartley y Rao (1967). En el análisis de variancia la función de probabilidad de una variable aleatoria depende de que los parámetros de la distribución (media y variancia) sean conocidos. En otras palabras, si conocemos la media y variancia de una distribución, podemos determinar para cualquier valor de la variable aleatoria cuál es su probabilidad. En el análisis de ML el razonamiento es inverso. Conocido un vector de observaciones, se trata de inferir a qué distribución de probabilidad responde. Es decir, tratamos de determinar de qué población o distribución de probabilidades de la variable aleatoria es más probable (o verosímil) que se haya obtenido el vector de observaciones.

8.10.1. Función de máxima verosimilitud

Esta función de densidad de probabilidad se identifica como si fuera una probabilidad condicional:

$$l(\theta|z)$$

es decir, la verosimilitud de obtener un vector de parámetros poblacionales (θ) dado el vector de observaciones (z). Los estimadores de máxima verosimilitud (MLE) de los parámetros desconocidos θ son los valores más probables para producir el vector de observaciones z .

Para determinar más fácilmente al estimador, se busca el máximo para la función de verosimilitud trabajando con logaritmos. De tal modo:


$$l(\theta|z) = \ln [l(\theta|z)]$$

Cuando se trabaja con muestras poblacionales de tamaño grande, los MLE tienen las siguientes propiedades:

- Son consistentes (a mayor tamaño de muestra, el estimador de θ tiende a converger con el valor real del parámetro).
- Son insesgados (invariantes).
- Poseen normalidad asintótica (a mayor tamaño de muestra, la distribución de los MLE converge a una distribución normal y ninguna otra distribución posee menor variancia que esta. Por lo tanto los MLE poseen menores intervalos de confianza).
- La variancia de un MLE es aproximadamente el recíproco de la segunda derivada de la función logaritmo de la verosimilitud con signo negativo.
- Son eficientes porque hacen uso de toda la información disponible y tienen en cuenta la eventual falta de independencia (presencia de correlación) entre las observaciones.

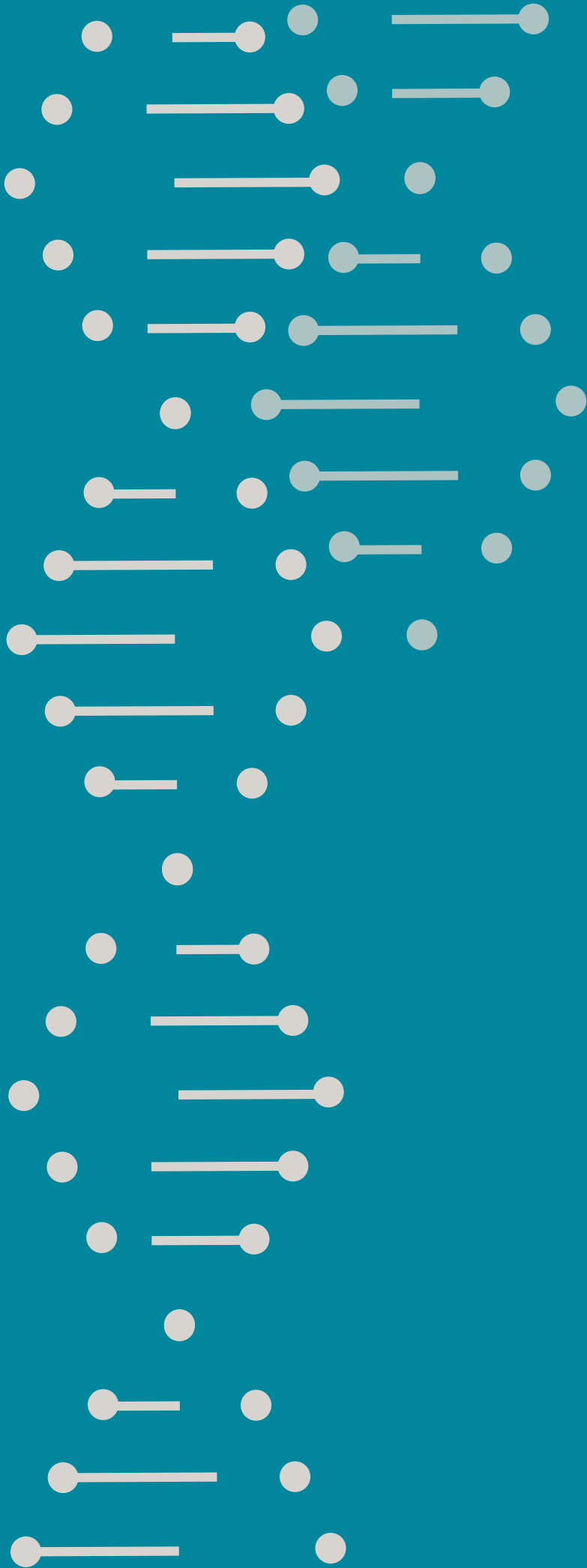
8.10.1.1. Algunas limitaciones del análisis de máxima verosimilitud

El análisis de ML asume que los efectos fijos se conocen sin error. Ello implica una limitación en el análisis ya que ese supuesto es de difícil cumplimiento. Consecuentemente las estimaciones de los componentes de variancias resultan sesgadas y la variabilidad no contemplada por el modelo (variancia residual) subestimada. Estas limitaciones se hacen más importantes cuanto mayor sea el número de factores fijos incluidos en el modelo y más pequeña sea la muestra poblacional. Las estimaciones con REML tienen la ventaja de maximizar la fracción de la probabilidad que no depende de los efectos fijos.



Estos inconvenientes del modelo ML, en lo que respecta al sesgo (subestimación de las variancias paramétricas), pueden morigerarse (aunque no eliminarse por completo) mediante el modelo REML (máxima verosimilitud restringida o máxima verosimilitud residual (Patterson y Thompson, 1971)).

El análisis REML es uno de los métodos más empleados para la estimación de parámetros en la mayoría de los diseños de genética cuantitativa. La resolución de los sistemas de ecuaciones REML (y ML) se obtiene a partir de procesos iterativos basados en diferentes algoritmos. Uno de los algoritmos más utilizados es el de Newton-Raphson y el EM. El punto de partida de las iteraciones es una estimación a partir de mínimos cuadrados ordinarios y la aplicación posterior del algoritmo en la dirección que maximice la verosimilitud.





Selección

La selección es un proceso sistemático que implica la existencia de una tasa de reproducción diferencial entre los individuos de una población que conduce a la modificación de sus frecuencias génicas. Esta tasa de reproducción diferencial puede obedecer a causas naturales o resultar de una decisión deliberada del fitomejorador en función de la constitución genética presunta de los individuos que selecciona de la población. En el primer caso nos referimos a la selección natural y en el segundo a la selección artificial.

La tasa de reproducción diferencial entre individuos significa que la cantidad de descendientes con los que contribuyen esos individuos al *pool* genético de la siguiente generación es diferente. La selección natural puede ocurrir

tanto en la etapa gametofítica (n) como en la esporofítica de los individuos ($2n$). En el primer caso la selección natural opera sobre las diferencias de fertilidad entre individuos. Puede suceder que se den situaciones de incompatibilidad gametofítica, impedimentos a la fertilización, o variabilidad en la aptitud de supervivencia de las gametas. Una vez generadas las cigotas o individuos que compondrán la generación siguiente, pueden darse diferencias de viabilidad entre ellas y en consecuencia ciertos esporofitos tendrán mayor probabilidad que otros de prosperar en su hábitat y llegar a la etapa reproductiva, generar gametas y dejar descendencia.

Desde el punto de vista del mejoramiento, en este texto nos dedicaremos fundamentalmente a la selección artificial, si bien es cierto que la selección natural puede afectar la respuesta a la primera y su existencia no puede descartarse por completo.

9.1. Tipos de selección

Dependiendo de la ubicación de los individuos seleccionados en la curva de distribución de frecuencias fenotípicas de la población pueden distinguirse tres tipos de selección. En la direccional, los individuos seleccionados se encuentran ubicados en una de las dos colas de la distribución. En la estabilizante, los individuos seleccionados son una fracción de los que se ubican a uno y otro lado de la media poblacional; y en la selección disrruptiva, los individuos seleccionados se ubican en ambos extremos o colas de la distribución (Figura 9.1).

En la medida que los tres tipos de selección sean efectivos y se sucedan varios ciclos de selección, se espera que ocurran cambios significativos en la media y variancia poblacional. En el caso de la selección direccional, la media poblacional experimentará un corrimiento en el mismo sentido de la selección. Es posible que la variancia poblacional se reduzca si los tamaños efectivos poblacionales son pequeños como para causar deriva genética. En el caso de la selección estabilizante es de esperar que con el transcurso de los ciclos de selección la distribución de frecuencias fenotípicas poblacionales no modifique su media, pero sí que se reduzca la variancia poblacional. Solo podría incrementarse la variancia poblacional si los genotipos heterocigotas tuvieran alguna ventaja adaptativa sobre los homocigotas. El mantenimiento de una presión selectiva de tipo disrruptivo puede provocar que la media poblacional no se altere, pero sí la forma típica de la distribución de frecuencias poblacional, generándose dos picos de frecuencia, cada una entre la ubicación de la media poblacional y las dos colas de la distribución.

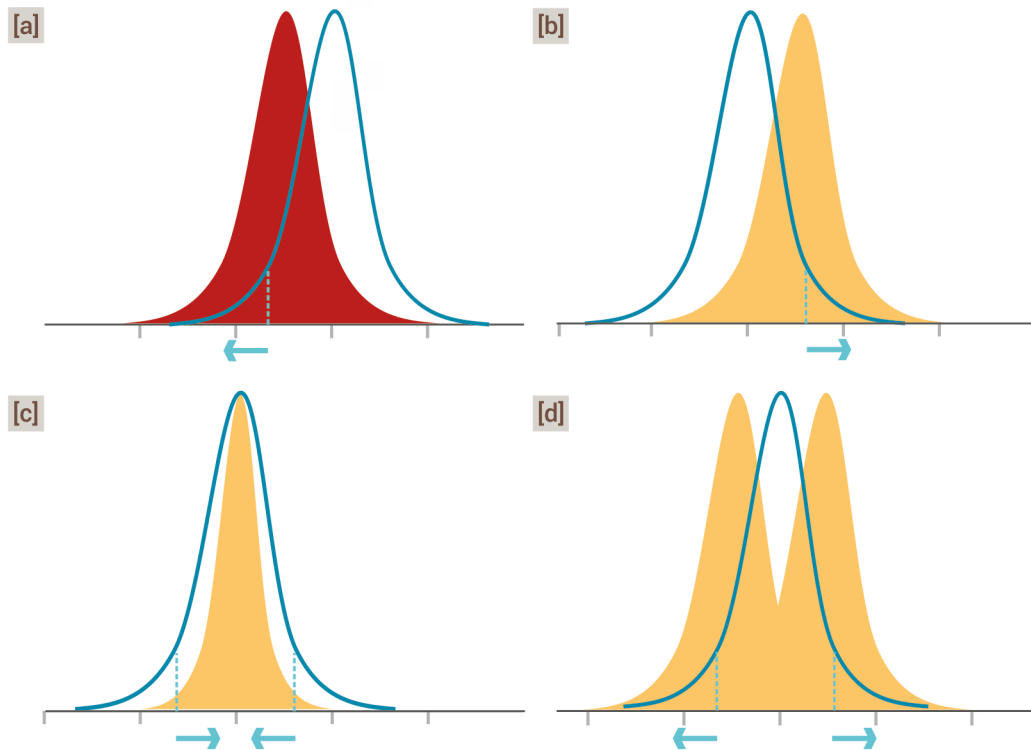


Figura 9.1. Tipos de selección. [a] Direccional en sentido negativo; [b] Direccional en sentido positivo, [c] estabilizante y [d] disruptiva. Las áreas coloreadas representan la distribución fenotípica esperable de la población luego de un número de ciclos de selección en la dirección indicada por las flechas.

9.2. Selección artificial

La selección artificial es la elección deliberada de un grupo de individuos para formar la generación siguiente. Generalmente se aplica en la etapa esporofítica. Es de práctica habitual que el fitomejorador defina un valor umbral para el carácter bajo selección y decida que ninguno de los individuos cuyos valores de mejora estimados sean inferiores a ese umbral contribuya con descendientes a la próxima generación. A este tipo de selección se lo denomina selección por truncamiento. Todos los individuos con estimaciones de valores de mejora ubicados por encima o por debajo (según la dirección de la selección) del punto de truncamiento serán los seleccionados y los únicos que contribuirán con sus genes a la siguiente generación en proporciones equivalentes unos de otros (Figura 9.2).

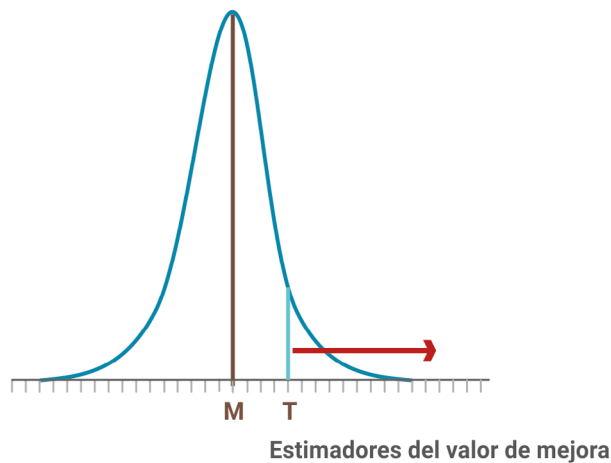


Figura 9.2. Punto de truncamiento (T) a partir del cual se seleccionan los individuos que poseen un valor fenotípico (o valor de mejora estimado) que lo iguala o supera. M representa la media poblacional original. El esquema corresponde a selección direccional a favor del carácter.

A menos que actúe la selección natural, las frecuencias génicas de los individuos seleccionados son las que determinarán las frecuencias génicas en la siguiente generación, es decir, en la población mejorada. La contribución gamética de individuos no seleccionados a la siguiente generación será nula en la medida que el control parental sea estricto (ver más adelante).

9.3. Algunas definiciones en relación con la selección artificial

Siguiendo a Comstock (1996), las unidades de selección son las entidades entre las cuales se practica selección (ejemplo: plantas individuales, familias de medios hermanos, etc.). Las selecciones son las unidades de selección específicamente elegidas como progenitores de la siguiente generación (ejemplos: planta individual A-104, planta individual A-219, familia 4, familia 99, familia 331, etc.). Se denomina criterio de selección al carácter o variable (o función de variables) sobre la cual se basa la selección, medido en una unidad de selección determinada (ejemplos: estimadores o predictores del valor de mejora para el tamaño relativo del embrión en la semilla de individuos, contenido de aceite de familias de medios hermanos, rendimiento de grano en familias de hermanos completos, requerimiento térmico de familias endocriadas, índice de selección basado en humedad de grano, sanidad foliar y rendimiento de familias de medios hermanos interpopulacionales, presencia/ausencia de un marcador molecular de QTL, en plantas individuales F_2 , etc.). Criterio de respuesta es la medida fenotípica con la que se expresa la respuesta a la selección. Cuando el criterio de respuesta y el criterio de selección coinciden nos estaremos refiriendo a la respuesta directa a la selección (ejemplo: cambios en la media de

rendimiento de familias de medios hermanos). Cuando los criterios de respuesta y de selección son diferentes nos referimos a respuesta indirecta a la selección (ejemplo: los cambios ocurridos en la tolerancia a la salinidad de familias de medios hermanos seleccionadas por rendimiento de grano en condiciones no limitantes).

9.4. Respuesta a la selección direccional

La selección direccional es el tipo de selección artificial típicamente aplicada sobre caracteres de interés económico, donde el objetivo es desplazar la media de la población en la dirección deseada. Los cambios en la media poblacional ocurrirán siempre que la selección direccional haya sido efectiva en aumentar las frecuencias de los alelos favorables para el carácter. En ausencia de efectos de la selección natural, hemos señalado que las frecuencias génicas en la población mejorada serán iguales a la frecuencia promedio de los individuos seleccionados. Desafortunadamente, cuando la selección es fenotípica, no hay manera de tener certeza sobre la constitución genética de las selecciones. Descansamos en el supuesto de que los individuos o familias que mostraron los valores fenotípicos superiores lo hacen por poseer un valor aditivo superior (o mayor valor reproductivo o de mejora), asumiendo que los eventuales efectos ambientales han sido razonablemente controlados y que la dominancia no es la acción génica que predomina para explicar la variabilidad fenotípica entre las unidades de selección para ese carácter.

Asumamos que tenemos una población con distribución fenotípica gaussiana y que una vez definido un punto de truncamiento quedan identificados los individuos de la población original (C_S) que aportarán sus gametas para producir los descendientes que compondrán la población mejorada (C_M). En un gráfico de coordenadas (x, y) (Figura 9.3) podemos ubicar como punto de origen (C_0) a la media de la población original.

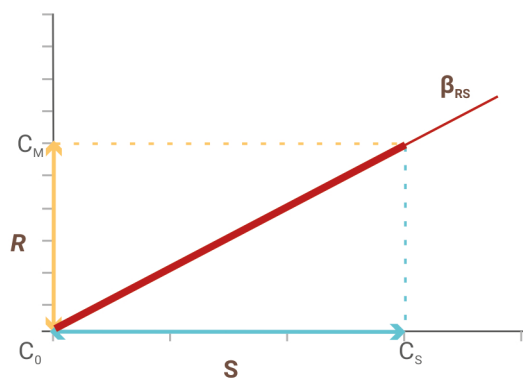



Figura 9.3. Representación del diferencial de selección S y la respuesta a la selección R . C_0 , C_S y C_M indican las medias de la población original, la media de las selecciones y la media de la población mejorada.

Sobre el eje de la abscisa se determina la media de las selecciones de la población original, indicada por C_S . El segmento C_0C_S representa el diferencial de selección S , es decir la diferencia entre la media fenotípica de la población original y la media de los individuos seleccionados. Sobre el eje de las ordenadas podemos ubicar un punto C_M correspondiente a la media de la población mejorada (C_M). La distancia o segmento C_0C_M representa la respuesta a la selección R , definida como la diferencia entre las medias de la población mejorada y la población original ($C_M - C_0$). Si repitiéramos el procedimiento un número suficiente de veces para distintos diferenciales de selección (segmento C_0C_S), seríamos capaces de trazar una línea de regresión que vincula la respuesta R (variable dependiente) con el diferencial de selección S (variable independiente). Esa línea de regresión tiene una pendiente β_{RS} que es la heredabilidad del carácter en sentido estricto. Esta pendiente tiene como valor máximo 1 y como valor mínimo 0. Si hubiese una correspondencia directa entre S y R ($h^2 = 1$), ello significaría que sería posible transferir íntegramente los valores fenotípicos de los individuos seleccionados a su descendencia. Ello es imposible de lograr por varias razones. En primer lugar, porque los valores fenotípicos de los individuos seleccionados no responden exclusivamente al componente aditivo, sino también al “no aditivo” y al desvío ambiental. Estos últimos no pueden capitalizarse por selección, por cuanto el componente ambiental es, por definición, no heredable y porque en especies diploides lo que se transmite de una generación a la siguiente no son las interacciones intraalélicas (dominancia) presentes en C_0 , sino los alelos con sus efectos medios. Por la misma razón tampoco se transmiten las interacciones interalélicas de tipo dominante x dominante y aditiva x dominante y solo parcialmente las de tipo aditivo x aditivo, en la medida que no sean interrumpidas por efecto de la recombinación genética (*crossing over*). En síntesis, la segregación mendeliana y la recombinación impiden la transferencia de los genotipos más destacados de C_0 a C_M . También puede suceder que algunos individuos hayan sido seleccionados por sus valores fenotípicos excepcionales resultado de haber estado expuestos a ambientes excepcionalmente favorables para el carácter seleccionado.

Geoméricamente llegamos a la siguiente expresión que vincula la respuesta a la selección con el comportamiento promedio de los progenitores y la heredabilidad (o pendiente de la recta de regresión β_{RS}):

$$R = S h^2, \quad [9.1]$$

A partir del conocimiento que tengamos sobre la heredabilidad y del comportamiento promedio de los individuos seleccionados en C_0 es posible predecir la respuesta esperada a la selección artificial.



Como veremos más adelante, el numerador que se emplee en el cómputo de la heredabilidad dependerá del tipo de selección que se realice y de los genotipos que se evalúen. Cuando el objetivo es seleccionar los mejores individuos de una población de cría de una especie de reproducción agámica (es decir, que cada individuo puede multiplicarse asexualmente sin cambio genotípico), o el mejor híbrido F_1 de una especie alógama, o la mejor línea pura de una especie autógena, podemos capitalizar todos los tipos de acción génica y por lo tanto utilizar la heredabilidad en sentido amplio en la fórmula de respuesta a la selección. En realidad, en el caso de líneas de especies autógenas, solo existirán como componentes de la variancia genotípica aquellos que involucren exclusivamente la acción aditiva (variancia aditiva y los tipos de variancia epistática aditiva x aditiva, aditiva x aditiva x aditiva, etc.) (Dudley y Moll, 1969). Si queremos mejorar una población de cría seleccionando individuos directamente (*per se*) o con base en pruebas de su progenie (familias), el numerador de la heredabilidad será el componente de variancia aditiva que se exprese entre esas unidades de selección (o tipo de familias).

El comportamiento promedio de la población luego de la selección es una consecuencia de las decisiones del fitomejorador, especialmente las que toma para modificar (aumentar) la heredabilidad del carácter.

Como se indicó más arriba, la selección direccional es efectiva si ha provocado incrementos en la frecuencia de los alelos favorables. Como sabemos, tales cambios modifican a su vez las variancias de origen genético y a sus proporciones relativas, por lo que la estimación de la heredabilidad que se utiliza para predecir la respuesta va perdiendo confiabilidad con la sucesión de ciclos de selección. Como los cambios en las frecuencias alélicas causadas por cada ciclo de selección son muy pequeños y a los fines prácticos, la estimación disponible de heredabilidad puede utilizarse por un número reducido de ciclos y luego reestimarse.

9.4.1. Diferencial de selección e intensidad de selección

Las distribuciones fenotípicas en poblaciones naturales para la mayoría de los caracteres de interés tienden a ajustarse a una distribución normal. Al comparar poblaciones generalmente se observa que poseen variancias diferentes. Asumamos que existen tres poblaciones bajo selección, A, B y C, que poseen la misma media, que la variancia fenotípica de la población A es menor que la de la población B y la de esta menor que la de C y que cada una de ellas se selecciona en forma direccional utilizando un mismo punto de truncamiento, T. En estas condiciones, si T lo expresamos en las mismas unidades de medida que el carácter seleccionado, quedarán a la derecha de T porcentajes de familias o individuos para seleccionar diferen-

tes dependiendo de la variancia fenotípica de cada población. En el caso de la población con menor variancia, el porcentaje de familias seleccionadas será menor que en la población con mayor variancia, dicho de otra manera, la selección será más exigente (Figura 9.4).

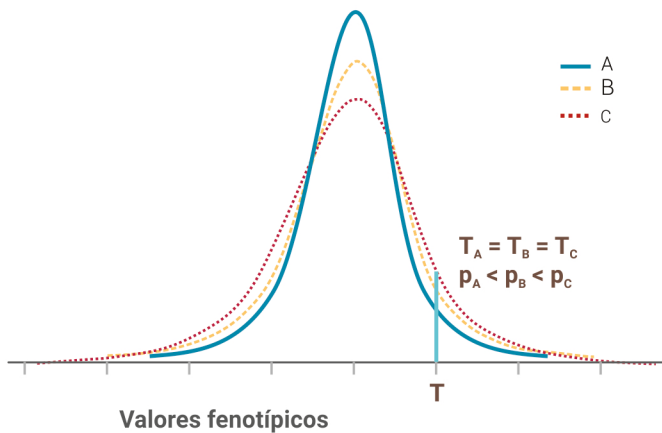


Figura 9.4.

Efecto de diferentes variancias de los valores fenotípicos de tres poblaciones de igual media respecto de la ubicación del punto de truncamiento medido en las unidades métricas del carácter seleccionado y su relación con el porcentaje de individuos seleccionados (p_A , p_B , p_C ; área de la distribución a la derecha del punto de truncamiento ($T=T_A=T_B=T_C$)).

Para predecir la respuesta a la selección a partir del diferencial de selección, el fitomejorador deberá conocer cuál es la media de los individuos o familias seleccionadas y restarle la media de la totalidad de los individuos o familias evaluadas, o media de la población.

La distribución normal, $N(0,1)$, posee algunas características de interés que, aplicadas a poblaciones naturales que siguen esta distribución, servirán para predecir la respuesta a la selección con solo conocer el porcentaje de familias o individuos seleccionados respecto del total evaluado. Si los puntos de truncamiento para las tres poblaciones A, B y C fueran iguales, pero no definidos en las unidades de medida de la característica fenotípica, sino que se expresaran como su ubicación en número de desvíos estándares de la población respectiva por encima de la media ($x\sigma_i$), los porcentajes de individuos o familias seleccionadas en cada una de las poblaciones serán idénticos. En la medida que las poblaciones se ajusten a una distribución normal, habrá una correspondencia entre la proporción de individuos o familias seleccionadas y el diferencial de selección medido en unidades de desvío estándar (o intensidad de selección). En tal caso el fitomejorador puede predecir las respuestas esperadas utilizando distintas intensidades de selección, simplemente sobre la base de la proporción de individuos o familias seleccionadas, evitando tener que calcularlo en cada caso a partir de las medias observadas de los individuos seleccionados (Figuras 9.5 y 9.6).

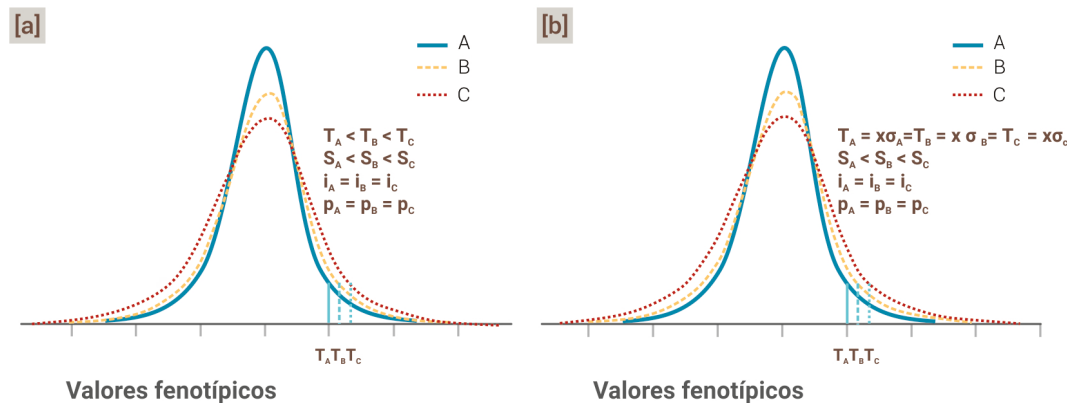


Figura 9.5. Efecto de diferentes variancias de los valores fenotípicos de tres poblaciones. [a]: para una misma proporción de individuos seleccionados ($p_A = p_B = p_C$), corresponden diferentes puntos de truncamiento ($T_A \neq T_B \neq T_C$) y diferentes diferenciales de selección ($S_A < S_B < S_C$). [b] para un mismo punto de truncamiento medido en desvíos estándar del carácter bajo selección, corresponderá igual proporción de individuos seleccionados e iguales diferenciales de selección medidos en desvíos estándar (i) y diferentes diferenciales de selección en las unidades métricas específicas al carácter.

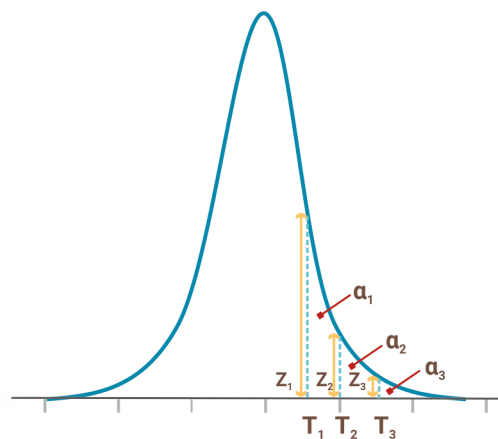


Figura 9.6. Relación entre el punto de truncamiento (T_1 a T_3), la ordenada en dicho punto (Z_1 , Z_2 , Z_3) y la proporción de individuos seleccionados (α_1 , α_2 , α_3) en una población con distribución normal para el carácter de interés.

Bajo la distribución normal existe una relación entre la ordenada en el punto de truncamiento (z_T) y la probabilidad o área bajo la curva a la derecha de aquel punto de truncamiento (α):

$$i = z_T / \alpha,$$

donde i es la intensidad de selección.

Basado en la función de densidad de probabilidad y función de densidad de probabilidad acumulada es posible que, conocida la intensidad de selección, pueda calcularse la media esperada de los individuos seleccionados (Figura 9.7).

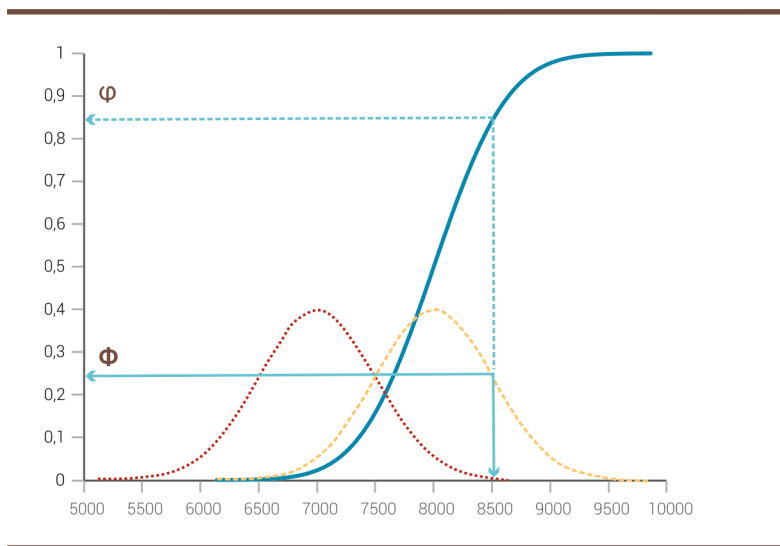


Figura 9.7. Funciones de densidad de probabilidad y de probabilidad acumulada para diferentes poblaciones. En una población con media 8000 y desvío estándar de 500, coincidente con el punto de truncamiento $T=8500$, pueden determinarse las ordenadas al origen Φ y ϕ . Basado en Xavier *et al.*, 2016.

La media esperada de los individuos seleccionados (y del ciclo siguiente asumiendo que la heredabilidad fuera igual a 1) será (Xavier *et al.*, 2016):

$$M_{C1} = M_{C0} + \sigma[\phi/(1 - \Phi)]$$

En el caso de la Figura 9.7, la media esperada será:

$$M_{C1} = 8000 + 500[0,85/(1 - 0,78)]$$

Cuanto mayor sea el porcentaje de individuos (o familias) seleccionado, menor será el diferencial de selección expresado en desvíos estándar. El diferencial de selección así expresado se denomina intensidad de selección (i). Entonces, cualquiera sea la población, siempre que la distribución sea normal, podremos predecir la respuesta a la selección a partir del conocimiento de la heredabilidad y del porcentaje de familias o individuos seleccionados. Con este último dato y mediante la utilización de tablas al efecto, se puede estimar la intensidad de selección.

Como dijimos que la relación entre el porcentaje de selecciones y la intensidad de selección es inversa, cuanto menor porcentaje de familias se seleccionen, la media de estas estará ubicada a mayor distancia de la media poblacional medida en unidades estándar. En este sentido el punto de truncamiento T puede expresarse como:

$$T = M_0 + i(V_{P0})^{1/2} \text{ (Xavier et al., 2016),}$$

donde M_0 es la estimación de la media poblacional (μ_0) y $(V_{P0})^{1/2}$ es la estimación del desvío estándar fenotípico de la población (σ_{P0}).

Falconer y Mackay (2001) han provisto tablas para determinar la intensidad de selección para diferentes porcentajes de individuos seleccionados en forma direccional, bajo dos situaciones de tamaño poblacional (poblaciones de tamaño infinito y de tamaño finito). Un extracto de esta para las aplicaciones más comunes se presenta en las Tablas 9.1 y 9.2.

Porcentaje seleccionado	x	Intensidad de selección (i)	Porcentaje seleccionado	x	Intensidad de selección (i)
1	2,326	2,665	15	1,036	1,554
2	2,054	2,421	20	0,842	1,400
3	1,881	2,268	25	0,674	1,271
4	1,751	2,154	30	0,524	1,159
5	1,645	2,063	35	0,385	1,058
10	1,282	1,755	50	0,000	0,798

Tabla 9.1.

Intensidad de selección correspondiente a diferentes porcentajes de unidades de selección seleccionadas en una población de tamaño infinito. La columna x indica el desvío respecto a la media de la población medido en unidades estándar. Adaptado de Falconer y Mackay, 2001.

n	N; %											
	50	%	100	%	150	%	200	%	300	%	400	%
5	1,705	10,00	2,018	5,00	2,185	3,33	2,297	2,50	2,449	1,67	2,552	1,25
10	1,372	20,00	1,730	10,00	1,916	6,67	2,040	5,00	2,206	3,33	2,317	2,50
15	1,139	30,00	1,536	15,00	1,738	10,00	1,871	7,50	2,048	5,00	2,166	3,75
20	0,951	40,00	1,386	20,00	1,601	13,33	1,742	10,00	1,928	6,67	2,051	5,00
25			1,259	25,00	1,488	16,67	1,636	12,50	1,830	8,33	1,958	6,25

Tabla 9.2.

Intensidad de selección correspondiente a diferente número de unidades de selección seleccionadas (n) en una población de tamaño finito (N). Adaptado de Falconer y Mackay, 2001.

Conforme lo anterior, podemos modificar la ecuación de respuesta [9.1] empleando el concepto de intensidad de selección:

$$i = S/\sigma_p ;$$

$$R = h^2S = (\sigma_A^2 / \sigma_p^2) S = (\sigma_A \sigma_A / \sigma_p)(S/\sigma_p) = ih\sigma_A , \quad [9.2]$$

O bien:

$$h = \sigma_A / \sigma_P$$

$$R = h \sigma_A i = (\sigma_A / \sigma_P) \sigma_A i (\sigma_P / \sigma_P) = ih^2 \sigma_P, \quad [9.3]$$

9.4.1.1. Apartamiento de la distribución poblacional respecto de la distribución normal

Cuando se trata de un carácter cuantitativo es de esperar que la distribución de los valores genotípicos se ajuste a la distribución normal cuanto mayor sea el número de loci que lo controlan. Algunos de los factores que contribuirían al apartamiento de la normalidad son que las frecuencias genéticas fueran distintas a 0,5, que los efectos de los loci fueran desiguales entre sí, o presencia de acción génica dominante y depresión por endocría. La normalidad de ambas distribuciones de los valores genotípicos y fenotípicos depende de que la distribución de los desvíos ambientales también se distribuya según la curva normal. Por el teorema central del límite se espera que así se cumpla.

Las expresiones de la respuesta esperada a la selección [9.2] o [9.3] resultan correctas si asumimos que la curva de la distribución de los valores fenotípicos de los individuos se ajusta a la distribución normal. En la medida que existan crecientes apartamientos de esta condición, aquellas expresiones resultarán en aproximaciones decrecientemente confiables. Recordemos que los valores fenotípicos responden a la suma de al menos dos términos, los valores genotípicos y los desvíos ambientales, para los cuales también estamos suponiendo que se ajustan a sendas curvas de distribución normal (Figura 9.8).

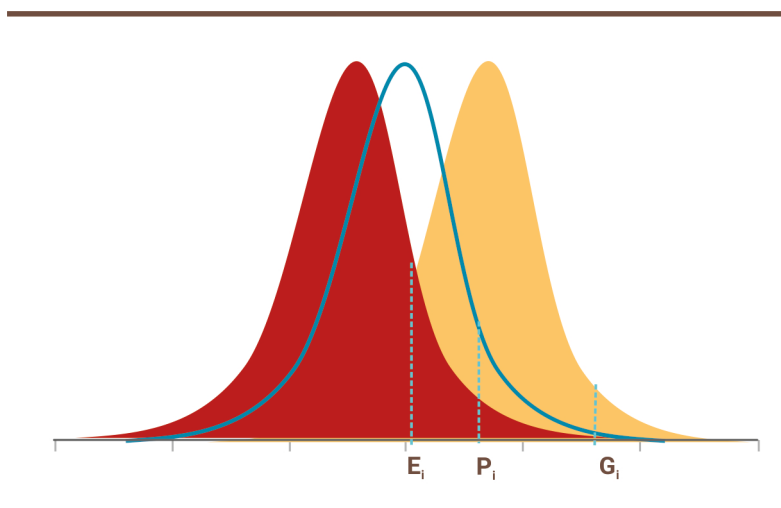



Figura 9.8. Ubicación hipotética del individuo i -ésimo dentro de la distribución de valores fenotípicos (P_i), genotípicos (G_i) y desvíos ambientales (E_i), asumiendo para todos ellos distribuciones normales.



Kelker y Kelker (1986), citando a Cochran (1951) y a Finney (1962) indican que cualquier población que inicialmente tuviera distribución normal, luego de un ciclo de selección direccional a favor del carácter resultará positivamente sesgada (ver Anexo). Con el transcurso de ciclos sucesivos de selección, este sesgo acumulado en la distribución poblacional puede tener un efecto importante sobre la calidad de la predicción de la respuesta. Mediante simulaciones pudo comprobarse que el sesgo aumentará cuanto mayor sea la heredabilidad y disminuirá con mayores fracciones de individuos seleccionados (menores intensidades de selección) (Ronningen, 1976). Empleando una metodología similar, Kelker y Kelker (1986) demostraron que la respuesta a la selección direccional a favor del carácter resultará menor en una población que tiene sesgo positivo respecto de otra con distribución simétrica, mientras que cuando la selección es en contra del carácter, la respuesta en la población sesgada será mayor que en la población con distribución simétrica. El apartamiento respecto a la distribución normal de los componentes genotípico y ambiental del valor fenotípico puede explicar la existencia de respuestas asimétricas cuando la selección es bidireccional (la misma población de partida se somete a dos selecciones, a favor y en contra del carácter).

Obsérvese que en especies de reproducción sexual, solo el componente aditivo de la variancia genotípica puede capitalizarse en la selección. Ello se debe a que en la etapa de intercrucamiento de las selecciones, la meiosis y recombinación genética no permiten la transferencia de interacciones entre alelos de un mismo locus (porque las gametas solo contienen un alelo de cada locus) ni podemos asegurar la transferencia de interacciones epistáticas favorables para el carácter entre alelos ubicados en diferentes loci debido al *crossing over*.

Es importante notar que las variancias aditiva y fenotípica se refieren a las unidades de selección que están siendo evaluadas. Tales unidades pueden ser simplemente individuos de la población C_0 que se evalúan *per se*, o bien familias derivadas de esos individuos que permiten hacerles una prueba de progenie en experimentos con repeticiones y, de ser posible, replicados en más de un ambiente. En este momento resulta oportuno referirnos a otra definición de la heredabilidad más ajustada a su aplicación práctica en programas de selección. En este sentido Hanson (1963) define a la heredabilidad como la proporción del diferencial de selección que puede capitalizarse en respuesta cuando la selección se practica sobre determinada unidad de selección.

9.5. Cambios de frecuencia génica por ciclo de selección

El cambio esperado en la frecuencia génica del alelo favorable por ciclo de selección, en ausencia de dominancia (Vencovsky, 1980), es:

$$\Delta_p = \frac{1}{2} [(sp_0q_0)/(1 - sq_0)], \quad [9.4]$$

donde s es el coeficiente de selección y p_0, q_0 son, respectivamente, las frecuencias del alelo favorable y desfavorable, previas a la selección.

Recordemos que el coeficiente de selección (s) es la disminución de la contribución gamética de un individuo a la siguiente generación en relación con la contribución gamética del genotipo más favorable. Se trata de una medida relativa de la presión de selección que experimenta un genotipo. En una población sometida a selección direccional por truncamiento, el coeficiente de selección de los genotipos no seleccionados será igual a cero.

Luego de t ciclos de selección (Wricke y Weber, 1986), el cambio en la frecuencia del alelo favorable será:

$$\Delta_p = p_t - p_0 = \sum_{j=0}^t sp_0q_0 [1 - (1/2N)]^j \approx 2Nsp_0q_0(1 - e^{-t/2N}), \quad [9.5]$$

donde p_t es la frecuencia alélica en la población luego de t ciclos de selección, y N es el número de genotipos seleccionados.

Falconer y Mackay (2001) presentaron, para el caso de un locus, la expresión aproximada de la equivalencia entre el coeficiente de selección (s) y el diferencial de selección (S):

$$s = 2ia/\sigma_p = 2Sa/\sigma_p^2$$

donde a es el desvío del homocigota favorable respecto del promedio de los dos homocigotas.

Reemplazando en [9.5]:

$$\Delta_p \approx 4NaSp_0q_0(1 - e^{-t/2N})/\sigma_p^2$$

Otra manera de analizar el cambio de frecuencia génica es partiendo directamente del diferencial de selección S (Wricke y Weber, 1986). Podemos asumir que hay una relación lineal entre el cambio de frecuencia génica y el incremento del diferencial de selección, de modo tal que:

$$\Delta_q = \beta_{qp}S, \quad [9.7]$$

donde β_{qp} es el coeficiente de regresión de la frecuencia génica en función del valor fenotípico. Como todo coeficiente de regresión, será la relación entre la covariancia entre las dos variables (frecuencia génica y valor fenotípico) y la variancia de la variable independiente (variancia fenotípica). Así:

$$\beta_{qp} = \text{Cov}(q, P) / \sigma_p^2$$

Recordemos la composición del valor fenotípico, de modo que

$$\beta_{qp} = \text{Cov}(q, G+E) / \sigma_p^2 = \text{Cov}(q, G) / \sigma_p^2$$

La covariancia entre la frecuencia génica del alelo favorable y el valor genotípico será (Tabla 9.3):

Genotipo	Frecuencia del alelo q	Valor genotípico	Frecuencia de q x Valor Genotípico
TT	0,0	a	0
Tt	0,5	d	½ d
Tt	1,0	-a	-a
Sumatoria	q	(1-2q)a + 2q(1-q)d	-q²a + q(1-q)d

Tabla 9.3.
Covariancia entre la frecuencia génica y el valor genotípico.

$$\text{Cov}(q, G) = [(1 - 2q)a + 2q(1 - q)d][-q^2a + q(1 - q)d] = -qp\alpha$$

$$\beta_{qp} = -qp\alpha / \sigma_p^2 \text{ y reemplazando en [9.6]}$$

$$\Delta_q = [-qp\alpha / \sigma_p^2] S$$

Reemplazando S por $i(\sigma_p^2)^{0,5}$,

$$\Delta_q = [-qp\alpha / \sigma_p^2] i \sigma_p = i[-qp\alpha / \sigma_p] , \quad [9.7]$$

Si asumimos que existen n loci que controlan el carácter y que idealmente toda la variancia fenotípica fuera aditiva ($\sigma_p^2 = \sigma_A^2$), entonces

$$\sigma_A^2 = n2qp\alpha^2$$

Por lo tanto el cambio máximo en la frecuencia génica por cada ciclo de selección será:

$$\Delta_q = -i qp \alpha / [n2qp\alpha^2]^{0,5}$$

$$\Delta_q = -ipq/[n2pq]^{0,5}$$

Cuando la selección es posterior a la floración, únicamente puede seleccionarse sobre las hembras, se reduce entonces el control parental y el cambio de la frecuencia génica es la mitad de la anterior [9.7]:

$$\Delta_q = -i[pq\alpha/2\sigma_p] , \quad [9.8]$$

Del análisis de la ecuación se desprende que el cambio de frecuencia génica por ciclo de selección será mayor cuanto mayor sea la intensidad de selección y cuanto menor sea la variancia fenotípica. Ambos factores son controlables por el fitomejorador.

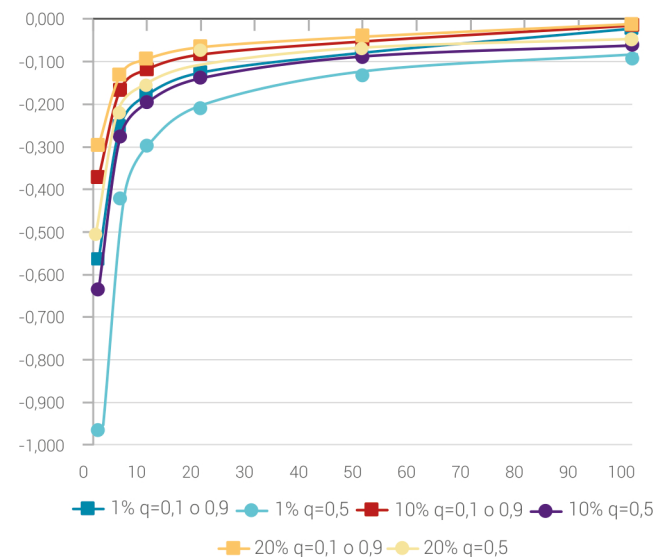


Figura 9.9. Cambios (disminución) en la frecuencia del alelo desfavorable por efecto de la selección para un carácter sin dominancia, controlado por un número variable de loci ($n=1, 5, 10, 20, 50, 100$) y tres porcentajes de individuos seleccionados (1 %, 10 % y 20 %) en una población con diferentes frecuencias del alelo desfavorable ($q= 0,1; 0,5; 0,9$).

Podemos además expresar la respuesta a la selección (R) en función del cambio de las frecuencias génicas. Para ello recordemos que:

$$R = C_M - C_0$$

y que las medias de las poblaciones son:

$$C_M = a(p_M - q_M) + 2dp_Mq_M , \quad [9.9]$$

$$C_0 = a(p_0 - q_0) + 2dp_0q_0$$

Los cambios en la frecuencia génica son:

$$\Delta p = p_M - p_0 \text{ y } \Delta q = q_M - q_0,$$

por lo tanto:

$$p_M = \Delta_p + p_0 \text{ y } q_M = \Delta_q + q_0 \quad [9.10]$$

y

$$\Delta_p = -\Delta_q, \quad [9.11]$$

Reemplazando en [9.9]:

$$\begin{aligned} C_M &= a[(\Delta_p + p_0) - (\Delta_q + q_0)] + 2d(\Delta_p + p_0)(\Delta_q + q_0) \\ &= a[(\Delta_p + p_0 + \Delta_p - q_0)] + 2d(\Delta_p + p_0)(-\Delta_p + q_0) \\ &= a[(2\Delta_p + p_0 - q_0)] + 2d(-\Delta_p^2 + \Delta_p q_0 - p_0 \Delta_p + p_0 q_0) \\ &= a(2\Delta_p) + a(p_0 - q_0) - 2d(\Delta_p^2) + 2d\Delta_p q_0 - 2dp_0 \Delta_p + 2dp_0 q_0 \\ &= \mathbf{a(p_0 - q_0) + 2dp_0 q_0} + 2a\Delta_p - 2d(\Delta_p^2) + 2d\Delta_p q_0 - 2dp_0 \Delta_p \end{aligned}$$

Reemplazando los términos en *negrita cursiva* por C_0 :

$$\begin{aligned} C_M &= C_0 + 2\Delta_p (a - d\Delta_p + dq_0 - dp_0) = C_0 + 2\Delta_p [a - d(\Delta_p + q_0 - p_0)] \\ &= C_0 + 2\Delta_p [a + d(-\Delta_p - q_0 + p_0)] \end{aligned}$$

Reemplazando por el efecto de sustitución del gen en C_0 :

$$\begin{aligned} \alpha_0 &= a + d(p_0 - q_0): \\ C_M &= C_0 + 2\Delta_p [a + d(p_0 - q_0) - d\Delta_p] \\ &= C_0 + 2\Delta_p [\alpha_0 - d\Delta_p] \end{aligned}$$

Por lo tanto:

$$R = 2\Delta_p [\alpha_0 - d\Delta_p] \quad [9.12]$$

La respuesta a la selección dependerá entonces del cambio de frecuencia y del efecto de sustitución de un gen, que a su vez es función de las frecuencias génicas de la población bajo selección.

En caso de no existir dominancia, recordemos que:

$$\alpha_0 = a,$$

en consecuencia:

$$R = 2\Delta_p a$$

9.6. Respuesta a diferentes métodos de selección

Las expresiones de respuesta a la selección [9.2] y [9.3] asumen que se evalúan individuos de la población C_0 , de la cual solo una fracción de ellos se seleccionan y se intercrujan para formar la población mejorada (C_M).

Si bien es cierto que la selección artificial en última instancia se ejerce sobre los individuos de la población original (C_0), con frecuencia se realiza una evaluación o prueba de las progenies derivadas de esos individuos (unidades de evaluación) y en la etapa de intercruzamiento o recombinación para generar la población mejorada, pueden utilizarse como unidades de recombinación las mismas o diferentes progenies a las utilizadas como unidades de evaluación.

Intuitivamente puede entenderse que la respuesta a la selección será mayor cuanto mayor sea el emparentamiento o relación genética entre los individuos que se evalúan y seleccionan de la población original, por una parte y los individuos que compondrán la población mejorada, por el otro. Es decir que la respuesta dependerá de qué unidades de evaluación y recombinación se utilizan y de las características de la evaluación y de la recombinación. Aquel grado de emparentamiento al que hacemos referencia se mide, como hemos visto en el capítulo 6, mediante la covarianza genética entre los individuos que se evalúan en la población original (C_0) y los individuos que compondrán la población mejorada (C_M).

En otras palabras:

$$R = f(\text{Cov}_{G(C_0, C_M)})$$

Esquemáticamente pueden describirse diferentes situaciones que vinculan o conectan a los individuos de C_0 y de C_M (Vencovsky, 1980; Vencovsky y Barriga, 1992). Designaremos los siguientes términos:

- P es un individuo perteneciente a la población original.
- X es una unidad de evaluación de P.
- Y es un individuo en la población mejorada.

La Figura 9.10 esquematiza dos alternativas de selección de individuos, donde la unidad de evaluación son directamente los individuos (evaluación *per se*) ($P=X$) y los individuos de la población mejorada (Y) surgen de dos formas de recombinación: en la alternativa 9.10 [a], todos los individuos Y proceden de la fusión de gametas de pares de individuos C_0 que han sido seleccionados. En la alternativa 9.10 [b], los individuos Y resultan de la fusión de dos gametas, pero solamente una de ellas proviene de individuos de C_0 seleccionados. Estas dos situaciones sirven para introducir el concepto de

control parental (c). En el esquema [a], el 100 % de la información genética de los individuos Y que componen la población mejorada ha sido transmitida por individuos seleccionados de la población original. En el esquema [b], solo podemos asegurar que los individuos Y de la población mejorada poseen al menos 50 % de información genética proveniente de individuos P seleccionados, esto es la procedente de la planta madre seleccionada que les dio origen. No existe control sobre la procedencia de la gameta masculina, de modo tal que los individuos Y proceden de la formación de cigotas cuya gameta masculina más probablemente no provendrá de individuos de C_0 seleccionados. Dicho en otros términos, el control parental en el esquema [a] es $c=1$ y en el esquema [b] es $c=0,5$.

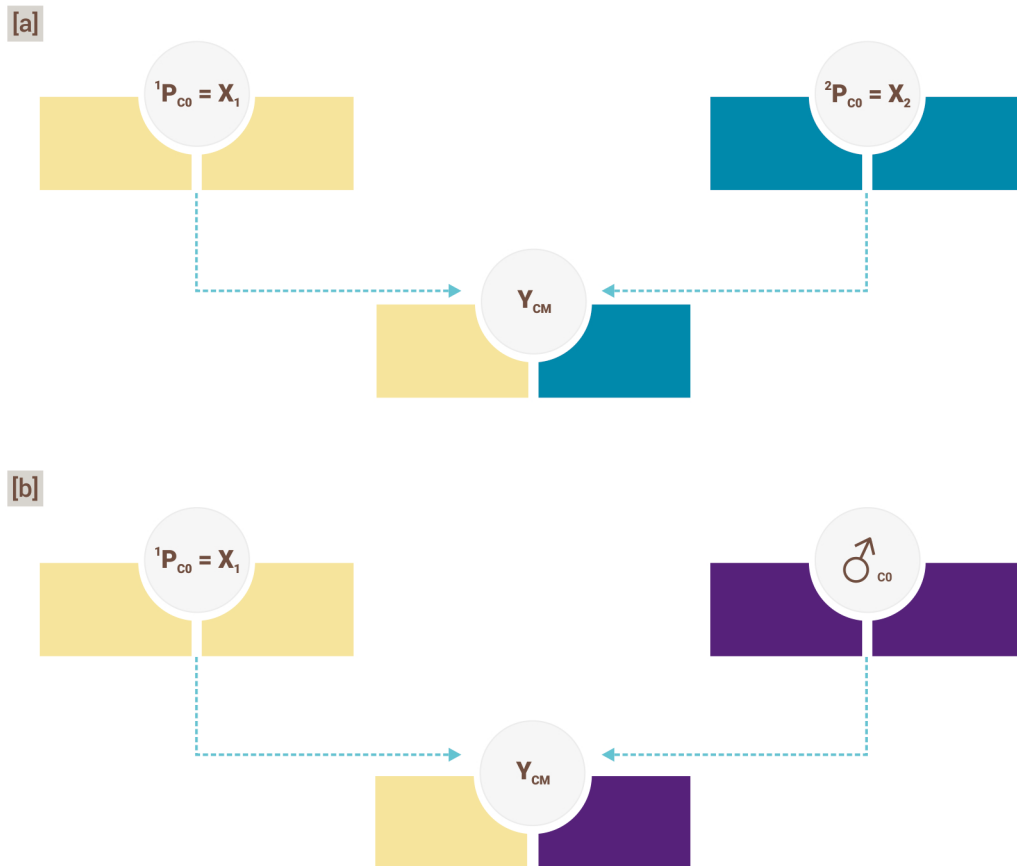


Figura 9.10.

Esquemas de relación entre individuos evaluados en la población bajo selección e individuos en la población mejorada. [a]: evaluación de individuos *per se* con recombinación solo entre individuos seleccionados. [b]: evaluación de individuos *per se* con recombinación entre individuos seleccionados y no seleccionados.

Comparativamente, la selección presentada mediante ambos esquemas (Figura 9.10), donde la evaluación de las unidades de selección se basa en el comportamiento *per se* de plantas individuales, no suele ser tan efectiva respecto a otros métodos, especialmente si el carácter de interés es de baja heredabilidad. Entre las razones para que ello ocurra podemos mencionar que al no disponerse de múltiples observaciones (repeticiones, ambientes) de cada individuo P de la C_0 , no es posible conocer o al menos estimar el error experimental ni la interacción genotipo x ambiente. En cambio, en especies de reproducción asexual o agámica, los individuos pueden reproducirse sin cambio genotípico y evaluárselos en experimentos con repeticiones y en diferentes ambientes.

Cuando el control parental es 0,5, la respuesta a la selección se reduce drásticamente al 50 %, por lo que la posibilidad de efectuar selección para el carácter antes de la floración es muy importante.

En la selección fenotípica de individuos *per se*, cada uno de ellos (P) es una unidad de selección y en consecuencia pueden utilizarse tamaños efectivos poblacionales bastante elevados y así aplicar altas intensidades de selección sin riesgo de que ocurra deriva genética.

Las expresiones de respuesta a la selección [9.2] y [9.3] son un caso especial en el que el control parental es 1. Las expresiones generales de respuesta considerando el control parental se modifican de la siguiente manera:

$$R = i c h \sigma_A = i c h^2 \sigma_P = i c \sigma_A^2 / \sigma_P, \quad [9.13]$$

Otro grupo de esquemas de selección que se presentan a continuación (Figuras 9.11 y 9.12) tienen en común la circunstancia de que las unidades de evaluación son familias derivadas de otros tantos individuos P de la población original C_0 y se incorporan unidades de recombinación R de cuyo intercrucamiento se obtendrá la población C_M . Las unidades de recombinación pueden estar constituidas por semilla remanente de la utilizada como unidad de evaluación, o puede ser un tipo de familia diferente.

De cada individuo P de C_0 se genera un tipo de familia X que se va a emplear para las evaluaciones. Parte de esta misma semilla, o la de otro tipo de familia originada de los mismos individuos P se mantendrá en reserva. Una vez que se haga la evaluación de las familias X y que se seleccionen los individuos P que dieron origen a las progenies X de mejor comportamiento, se acude a la semilla remanente de tales familias (R) que generarán los individuos Y de la población C_M .

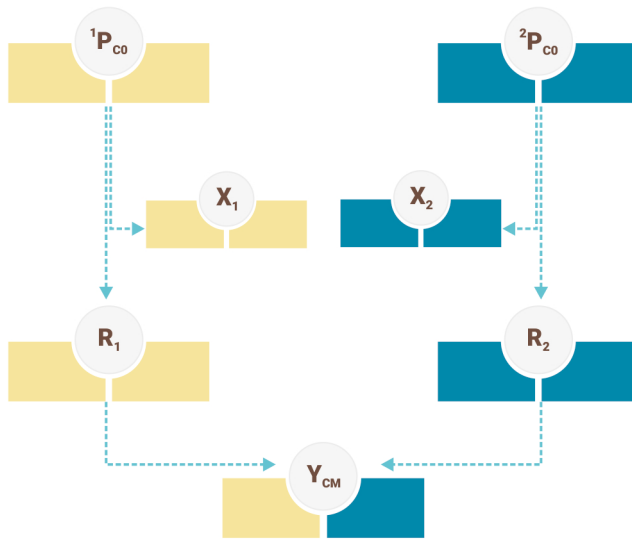


Figura 9.11. Esquema de selección donde los individuos P se evalúan por pruebas de progenie y la población mejorada se genera por recombinación entre unidades seleccionadas.

Los esquemas [a], [b] y [c] de la Figura 9.12 corresponden a situaciones donde el control parental es total ($c=1$). Es decir, que los individuos Y de C_M se formarán por fusión de gametas cuya trazabilidad solo llega a individuos P seleccionados. Los esquemas [d], [e] y [f] corresponden a situaciones donde el control parental es $c=0,5$. El símbolo ♂ representa muestras aleatorias de gametas provenientes de cualquiera de los individuos P de la población C_0 .

El esquema [a] representa la situación en la que las unidades de evaluación y de recombinación son semejantes y contienen información genética exclusivamente del individuo P de C_0 que les dio origen. En el esquema [b], cada unidad de evaluación (X) contiene 50 % de la información genética de los individuos P de la población C_0 y finalmente en el esquema [c], las unidades de evaluación (X) y las de recombinación (R) solo llevan la mitad de la información genética proveniente de los individuos P de C_0 . Los esquemas [d], [e] y [f] son respectivamente similares a los esquemas [a], [b] y [c], excepto por el hecho de que el control parental c es 0,5 debido a que los individuos Y solo tienen el 50 % de su información genética trazable a individuos P seleccionados.

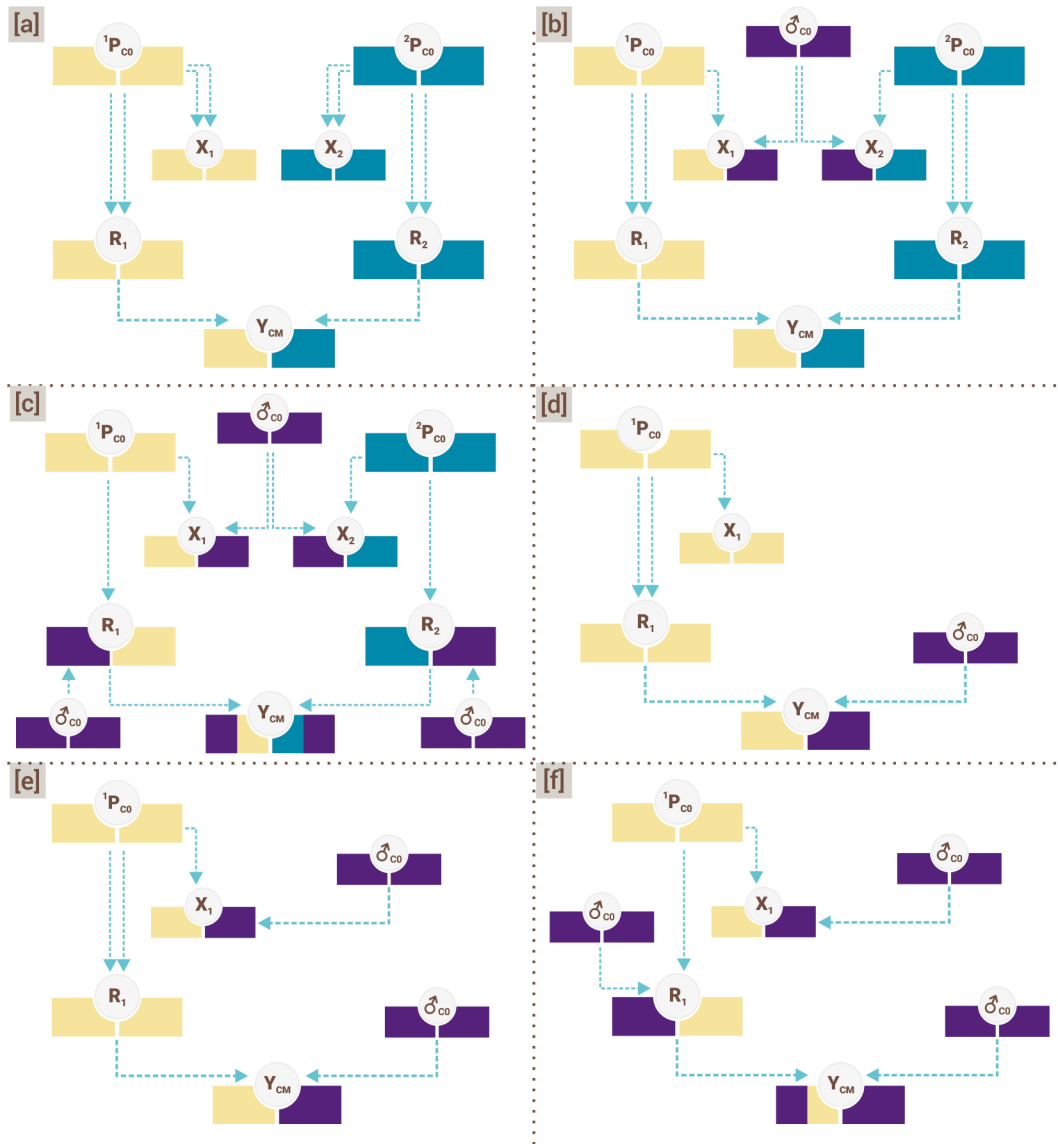


Figura 9.12. Diferentes esquemas de selección teniendo en cuenta variantes en cuanto al control parental, unidades de evaluación y de recombinación. Ver texto para las diferentes alternativas.

Si quisiéramos ordenar en forma decreciente la respuesta probable a los diferentes esquemas [a] a [f] en función de la magnitud de la covarianza genética entre las unidades de evaluación X, pasando por P, hasta llegar a un individuo Y en la población mejorada (C_M) resultarían respuestas con las siguientes características:

$$R_a > R_b > R_c \text{ y } R_d > R_e > R_f$$

Además, debido a las diferencias en el control parental:

$$R_a = 2R_d; R_b = 2R_e \text{ y } R_c = 2R_f$$

La respuesta a la selección para cualquier esquema selectivo puede predecirse utilizando la siguiente expresión general:

$$R = i \text{Cov}_{G(X,Y)} / (PV_X)^{0,5}, \quad [9.14]$$

donde $(PV_X)^{0,5}$ es el desvío estándar fenotípico de las unidades de evaluación.

A su vez, como hemos visto en el capítulo 7, la $\text{Cov}_{G(X,Y)}$ dependerá en primer lugar de cuánta variabilidad genética de la población queda expuesta entre las unidades de selección (que depende del tipo de progenies o familias que se están evaluando $(X, 2f_{xy} V_A + u_{xy} V_D)$ y del nivel de emparentamiento entre tales unidades de evaluación con el individuo Y en la población mejorada, que está afectado por el control parental:

$$\text{Cov}_{G(X,Y)} = c[2f_{xy} V_A + u_{xy} V_D]$$

Las diferentes respuestas en los esquemas [a] a [f] obedecen a que se modifica la magnitud de la covarianza genética entre las unidades de evaluación y los individuos de la población mejorada. El control parental afecta la respuesta al incidir sobre esa covarianza.

9.7. Factores que afectan la continuidad de la respuesta a través de ciclos de selección

La intensidad de selección y su efecto sobre el tamaño efectivo poblacional son los principales factores que afectan la persistencia de respuestas similares a lo largo de un número de ciclos de selección (Xavier *et al.*, 2016). El tamaño efectivo poblacional es el factor más limitante de la respuesta alcanzable por selección por cuanto si es pequeño puede provocar deriva genética y consecuente pérdida de variabilidad genética.

Los programas de mejoramiento pueden conducir la selección de sus poblaciones de cría con objetivos de largo y de corto plazo. En el primer caso se prioriza el compromiso entre magnitud de respuesta a la selección y mantenimiento de la variabilidad genética, de modo de sostener respuestas aceptables a lo largo de los ciclos de selección. En el segundo caso interesa obtener elevada tasa de respuesta por unidad de tiempo, sin importar el amesetamiento (*plateau*) de la respuesta luego de unos pocos ciclos de selección. El aporte de variabilidad desde germoplasma introducido o desde otras poblaciones a las que están siendo seleccionadas con objetivos

de corto plazo permite regenerar las condiciones que aseguren la recuperación del nivel de respuesta a la selección en estas últimas (Figura 9.13).

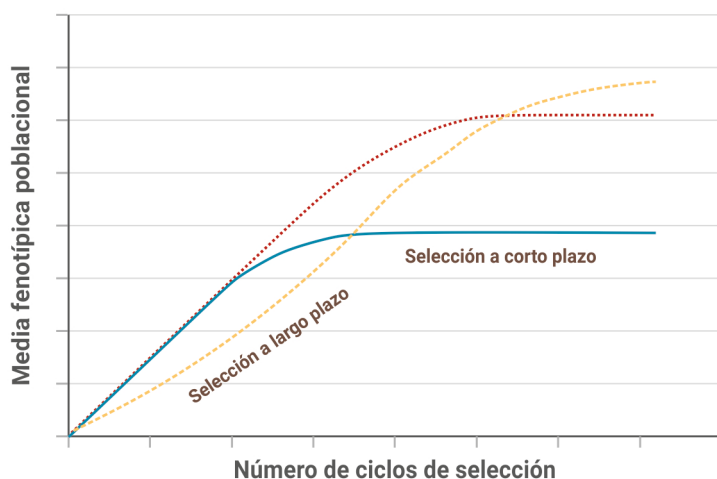


Figura 9.13. Estrategias de selección a corto y largo plazo. La línea roja punteada simula la superación del *plateau* en la estrategia de corto plazo mediante el aporte de variabilidad desde otras poblaciones o introducciones de germoplasma. Ver texto.

El énfasis relativo que se aplique a los objetivos de corto y largo plazo es característico de los programas de mejoramiento. En general, los programas públicos nacionales o internacionales priorizan objetivos de largo plazo, mientras que en los programas de mejoramiento comerciales predominan los objetivos de corto plazo. La necesidad de complementación entre ambos tipos de programas surge del beneficio para los programas de corto plazo del aporte de variabilidad genética novel proveniente de programas de largo plazo.

9.8. Selección entre y dentro de familias

Hemos visto que la variabilidad fenotípica en una población estructurada en familias puede subdividirse en sus componentes observacionales: la variancia fenotípica entre familias y la variancia fenotípica dentro de familias (ver Capítulo 5). Cada una de ellas puede, a su vez, desagregarse en los componentes causales de naturaleza genotípica y ambiental.

La distribución de la variancia genética o aditiva (componente fijo por selección) es diferente dependiendo del tipo de familia que se trate (Tabla 9.4).

Tipos de familias	Variancias entre familias			Variancias dentro de familias		
	Genotípica	Aditiva	Dominan.	Genotípica	Aditiva	Dominan.
MH	$(\frac{1}{4}) \sigma_A^2$	$(\frac{1}{4}) \sigma_A^2$	0	$(\frac{3}{4}) \sigma_A^2$	$(\frac{3}{4}) \sigma_A^2$	$1 \sigma_D^2$
HC	$(\frac{1}{2}) \sigma_A^2 + (\frac{1}{4}) \sigma_D^2$	$(\frac{1}{2}) \sigma_A^2$	$(\frac{1}{4}) \sigma_D^2$	$(\frac{1}{2}) \sigma_A^2 + (\frac{3}{4}) \sigma_D^2$	$(\frac{1}{2}) \sigma_A^2$	$(\frac{3}{4}) \sigma_D^2$
$S_{n-1:n}$	$2F \sigma_A^2 + (F/2^n) \sigma_D^2$	$2F \sigma_A^2$	$(F/2^n) \sigma_D^2$	$(1-F)[\sigma_A^2 + \sigma_D^2]$	$(1-F) \sigma_A^2$	$(1-F) \sigma_D^2$
$S_{0:1}$	$\sigma_A^2 + (\frac{1}{4}) \sigma_D^2$	σ_A^2	$(\frac{1}{4}) \sigma_D^2$	$(\frac{1}{2})[\sigma_A^2 + \sigma_D^2]$	$(\frac{1}{2}) \sigma_A^2$	$(\frac{1}{2}) \sigma_D^2$
$S_{1:2}$	$(\frac{3}{2}) \sigma_A^2 + (\frac{3}{16}) \sigma_D^2$	$(\frac{3}{2}) \sigma_A^2$	$(\frac{3}{16}) \sigma_D^2$	$(\frac{1}{4})[\sigma_A^2 + \sigma_D^2]$	$(\frac{1}{4}) \sigma_A^2$	$(\frac{1}{4}) \sigma_D^2$
$S_{2:3}$	$(\frac{7}{4}) \sigma_A^2 + (\frac{7}{64}) \sigma_D^2$	$(\frac{7}{4}) \sigma_A^2$	$(\frac{7}{64}) \sigma_D^2$	$(\frac{1}{8})[\sigma_A^2 + \sigma_D^2]$	$(\frac{1}{8}) \sigma_A^2$	$(\frac{1}{8}) \sigma_D^2$
$S_{3:4}$	$(\frac{15}{8}) \sigma_A^2 + (\frac{15}{256}) \sigma_D^2$	$(\frac{15}{8}) \sigma_A^2$	$(\frac{15}{256}) \sigma_D^2$	$(\frac{1}{16})[\sigma_A^2 + \sigma_D^2]$	$(\frac{1}{16}) \sigma_A^2$	$(\frac{1}{16}) \sigma_D^2$
$S_{1:3}$	$(\frac{3}{2}) \sigma_A^2 + (\frac{3}{16}) \sigma_D^2$	$(\frac{3}{2}) \sigma_A^2$	$(\frac{3}{16}) \sigma_D^2$	$(\frac{1}{4}) [\sigma_A^2 + \sigma_D^2]$	$(\frac{1}{4}) \sigma_A^2$	$(\frac{1}{4}) \sigma_D^2$
$S_{1:4}$	$(\frac{3}{2}) \sigma_A^2 + (\frac{3}{16}) \sigma_D^2$	$(\frac{3}{2}) \sigma_A^2$	$(\frac{3}{16}) \sigma_D^2$	$(\frac{1}{4}) [\sigma_A^2 + \sigma_D^2]$	$(\frac{1}{4}) \sigma_A^2$	$(\frac{1}{4}) \sigma_D^2$
$S_{2:4}$	$(\frac{7}{4}) \sigma_A^2 + (\frac{7}{64}) \sigma_D^2$	$(\frac{7}{4}) \sigma_A^2$	$(\frac{7}{64}) \sigma_D^2$	$(\frac{1}{8})[\sigma_A^2 + \sigma_D^2]$	$(\frac{1}{8}) \sigma_A^2$	$(\frac{1}{8}) \sigma_D^2$

Tabla 9.4.

Distribución de la variancia genotípica y sus componentes de aditividad y de dominancia para diferentes tipos de familias. MH y HC hace referencia a familias de medios hermanos y hermanos completos, respectivamente. S se refiere a familias de autofecundación ($S_{0:1}$, por ejemplo, se refiere a familias resultantes de la autofecundación de una planta de la población, que se asume sin endocria; $S_{1:2}$ se refiere a familias obtenidas luego de una generación de autofecundación de plantas individuales de familias $S_{0:1}$).

En el caso de las familias de medios hermanos (MH), el componente de la variancia aditiva es tres veces mayor dentro de familias que entre familias. En algunos métodos de selección recurrente, como el método espiga x hilera modificado (ver Capítulo 10), se aplica selección entre y dentro de familias de medios hermanos. La contribución de la variancia de dominancia a la variancia dentro de familias de hermanos completos (HC) es tres

veces mayor que la contribución a la variancia entre familias, mientras que la variancia aditiva se distribuye en las mismas proporciones (50 %) entre familias que dentro de familias. En familias de autofecundación conforme aumenta el nivel de endocría, va aumentando la variancia aditiva entre familias y va disminuyendo la variancia de dominancia, hasta llegar a cero cuando las familias alcanzan la homocigosis prácticamente completa. Dentro de las familias de autofecundación, tanto la variancia aditiva como la de dominancia expresan la misma fracción de las respectivas variancias poblacionales, pero ambas van disminuyendo hasta llegar a cero. En tal sentido, la selección dentro de familias autofecundadas producirá respuestas decrecientes a la selección cuanto mayor sea el nivel de endocría.

La selección entre líneas recombinantes o entre duplohaploides derivados de una población biparental tiene a disposición al doble de la variancia aditiva de la población de la cual derivan. En estos casos la variancia entre familias no resulta sesgada por componentes de dominancia, por cuanto en este tipo de poblaciones está garantizado que $p=q=0,5$ en todos los loci segregantes.

El conocimiento del patrón de distribución de los componentes genotípicos de la variabilidad resulta muy importante desde el punto de vista práctico para determinar la conveniencia y utilidad de la selección entre y dentro de familias (Tabla 9.4). Evidentemente la selección dentro de familias cobra sentido en la medida que la variabilidad dentro de ellas sea atribuible a la variancia aditiva, con poca incidencia de la dominancia y del ambiente. Es por ello que, en la conducción de filiales por el método genealógico, la selección dentro de familias no se justifica en etapas intermedias y avanzadas de endocría.

9.9. Repetibilidad de la respuesta a la selección Heredabilidad realizada

Si se llevaran a cabo en una población experimentos de selección y se repitieran empleando iguales intensidades de selección, será prácticamente imposible encontrar respuestas idénticas. Ello se debe a que las heredabilidades con la que se selecciona en uno u otro experimento muy probablemente serán diferentes debido a la imposibilidad de reproducir idénticas características ambientales en todos los experimentos.

La repetibilidad de la respuesta a la selección puede investigarse también desde otro punto de vista, no ya entre experimentos, sino analizando la respuesta observada a través de los ciclos del experimento de selección. La respuesta puede variar de un ciclo a otro por dos factores principales. Uno

de ellos ya fue mencionado: la variabilidad ambiental incontrolable que modifica la variabilidad fenotípica. El segundo son las modificaciones de las frecuencias génicas por efecto de los ciclos de selección precedentes y por lo tanto de la variancia genotípica y sus componentes. Ambos factores pueden afectar al denominador y numerador de la expresión que describe la heredabilidad.

Uno de los procedimientos para estimar la heredabilidad en una población para un carácter cuantitativo es conducir un experimento de selección durante una serie de ciclos y estimar el coeficiente de regresión entre la respuesta acumulada desde el C_0 al C_n , como variable dependiente, en función del diferencial de selección acumulado desde el C_0 al C_n , como variable independiente (Figura 9.14). El coeficiente de regresión o pendiente de la recta ajustada es por definición la heredabilidad realizada.

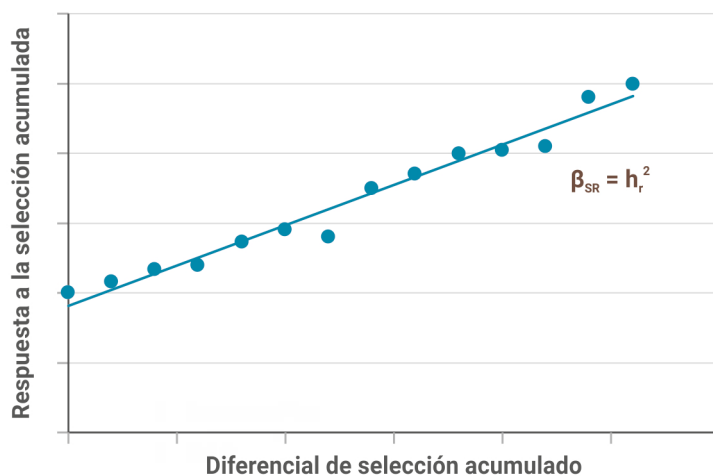


Figura 9.14. Respuesta acumulada a la selección en función del diferencial de selección acumulado, representado para un caso hipotético de quince ciclos de selección. La pendiente de la regresión (β_{SR}) es la heredabilidad realizada (h_r^2).

9.10. Respuesta asimétrica a la selección

Como se demostró anteriormente (Capítulo 5), las variancias aditiva y de dominancia (y la de epistasis) se modifican toda vez que cambien las frecuencias génicas. Según el modelo genético que se considere, se verificará un valor máximo de las variancias que ocurrirá a distintas frecuencias génicas del alelo favorable (Tabla 9.5).

Modelo genético		Frecuencia génica del alelo favorable (p) a la que la variancia es máxima		
		Aditiva	De dominancia	Genotípica
Aditividad completa	d=0	0,50		0,50
Dominancia completa	d=a	0,25	0,50	0,30
Sobredominancia	a=0	0,15 y 0,85	0,50	0,50

Tabla 9.5. Efecto de las frecuencias génicas sobre la variancia genotípica y su componente aditivo y de dominancia.

Aquellas variancias se hacen cero si las frecuencias génicas alcanzan los extremos ($p=0, q=1$; $p=1, q=0$). En el resto del intervalo (0,1), conforme aumentan (o disminuyen) las frecuencias alélicas, las variancias van modificándose. En el caso de un carácter sin dominancia, la variancia genética o aditiva será máxima cuando $p=q=0,5$ (Falconer y Mackay, 2001). Si la población original posee una frecuencia de alelos favorables dentro del intervalo (0 a 0,5), la selección direccional, en tanto sea efectiva, causará aumentos de la frecuencia génica provocando incrementos de la variancia aditiva y por lo tanto de la heredabilidad (a condición de similar contribución de la variancia ambiental a la variancia fenotípica). Se esperarían, dentro de aquel intervalo de frecuencias génicas, respuestas crecientes a la selección. Por el contrario, en el intervalo (0,5 a 1,0), la respuesta a la selección será decreciente a lo largo de los ciclos a medida que la frecuencia del alelo favorable se vaya aproximando a 1,0. Las mayores respuestas ocurrirán con frecuencias $p=q=0,5$ o cercanas a ese valor.

Cuando se realiza selección bidireccional bajo un modelo totalmente aditivo, la respuesta será simétrica (igual magnitud absoluta, pero con diferente signo) únicamente cuando $p=q=0,5$, situación que se denomina condición de simetría. A frecuencias diferentes, la respuesta a la selección en uno y otro sentido se espera que sean de distinta magnitud, porque las variancias aditivas y la heredabilidad serán diferentes en uno y otro sentido de la selección. Para un carácter que presenta dominancia completa, la condición de simetría se da cuando $p=0,25$ (Falconer y Mackay, 2001).

Desde el punto de vista práctico, para el fitomejorador el tema de la asimetría de la respuesta no resulta relevante en tanto y en cuanto normalmente le interesa mejorar la población en uno solo de los sentidos, pero sí es importante que se relacione la expectativa de respuesta a la selección con la respuesta observada por el efecto que tiene la ubicación de la frecuencia génica de los alelos favorables en la población de partida, al momento de iniciar la selección.

9.11. Límite a la selección

El límite teórico a la selección o respuesta teórica total máxima para un carácter se alcanza cuando los mejores alelos se fijan en cada uno de los loci que lo controlan. Si el alelo favorable está en muy baja frecuencia, aunque tuviera alguna ventaja selectiva, la selección tendrá poco efecto en su fijación a menos que el tamaño efectivo poblacional no sea muy pequeño (Figura 9.15). Recordemos que la probabilidad de fijación de un alelo por el solo efecto de la deriva genética es igual a su frecuencia.

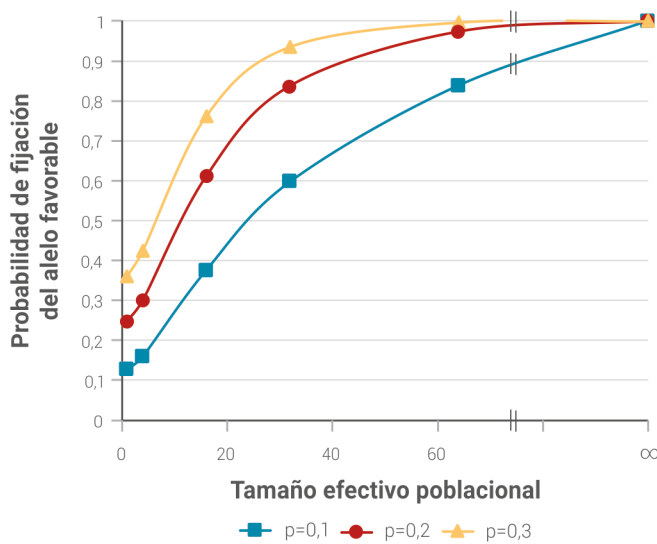


Figura 9.15. Probabilidad de fijación del alelo favorable por selección direccional asumiendo diferentes frecuencias génicas y diferentes tamaños efectivos poblacionales. A partir de Allison y Curnow, 1969.

Desde el punto de vista práctico, existen diferentes factores que inciden para que el límite real de selección se alcance antes que ese límite teórico. Bailey y Comstock (1976) señalan entre estos factores a los siguientes:

- Número de loci segregantes para el carácter y frecuencia alélicas en la población de partida.
- Magnitud de los efectos génicos.
- Tipos de acción génica.
- Tamaño efectivo poblacional.
- Intensidad de selección.
- Efectos pleiotrópicos y ligamiento.
- Mutaciones.
- Interacciones entre los factores mencionados más arriba.

Mediante un estudio de simulación Allison y Curnow (1969) estimaron las respuestas a la selección para un carácter aditivo, controlado por 150 loci y con un efecto $a=2$, utilizando diferentes tamaños efectivos poblacionales ($N_e = 1, 4, 16, 32, 64, 256$ e infinito). Concluyeron que el tamaño efectivo tiene un efecto muy importante sobre la respuesta máxima alcanzable luego de un número variable de ciclos de selección. El empleo de tamaños efectivos entre 16 y 64 serían los más convenientes desde un punto de vista práctico (Figura 9.16).

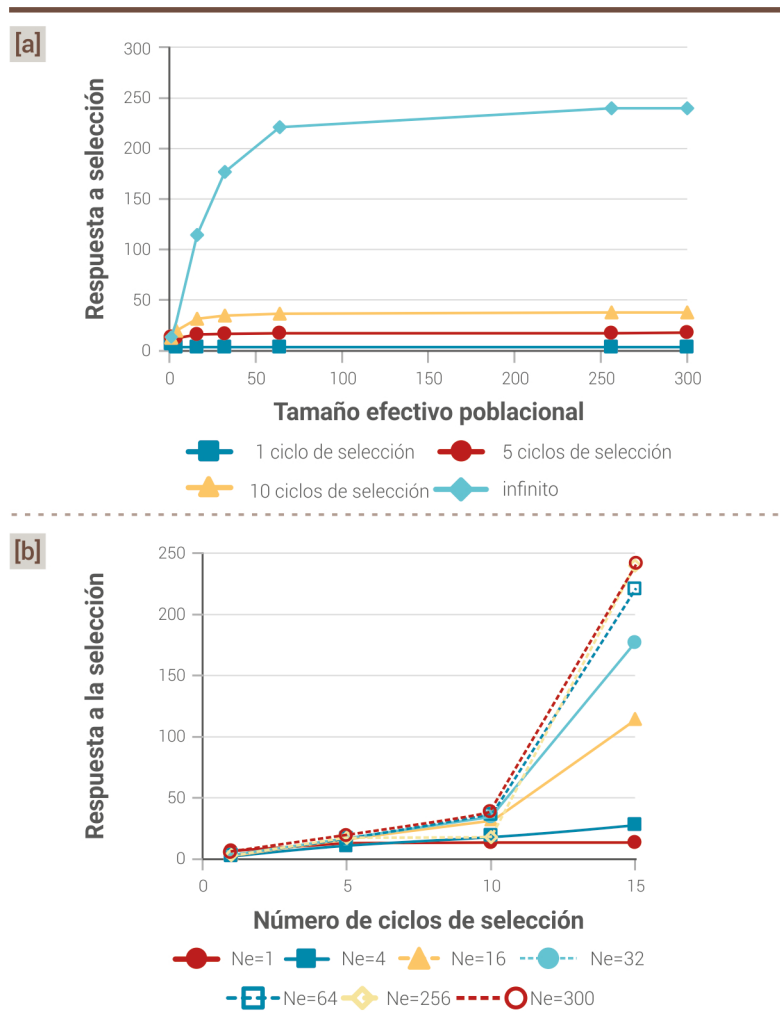


Figura 9.16. Efecto del tamaño efectivo poblacional sobre el límite alcanzable por selección. A partir de Allison y Curnow (1969), simulando una población de media inicial 60, máxima respuesta posible de 300 asumiendo la fijación del alelo favorable en 150 loci. [a] Respuesta esperada para diferentes tamaños efectivos poblacionales. [b] Respuesta esperada luego de un número variable de ciclos de selección.

Resulta de interés estimar el tiempo en que se alcanzaría el límite teórico y medirlo en número de ciclos de selección, en años o estaciones de crecimiento de cultivo. Para un carácter de herencia simple y selección en contra de un alelo dominante, el límite se alcanzará en apenas un ciclo de selección. En el caso de herencia simple y selección en contra de un alelo recesivo, el número de ciclos necesarios aumenta, debido a que los alelos recesivos que están enmascarados en los individuos heterocigotas no

son alcanzados por la selección. Para caracteres de herencia cuantitativa, cuanto mayor sea el número de loci que los controlan, mayor será el número de ciclos necesarios para alcanzar el límite de selección.

La respuesta a la selección no es necesariamente lineal a lo largo de los ciclos, sino que muestra una aproximación asintótica al límite de selección, hasta hacerse cero. La vida media de la selección (Robertson, 1960) es el tiempo (medido en número de ciclos) que se requiere para obtener el 50 % de la respuesta teórica total. Bajo un modelo genético sin dominancia, esto es equivalente a decir el tiempo necesario para lograr la mitad del avance en la frecuencia génica. Baker y Curnow (1969) demostraron que para el número de ciclos de selección que resulta usual en programas de mejoramiento, no se justificaría aumentar demasiado el tamaño efectivo por encima de $N_e=16$ (Figura 9.17).

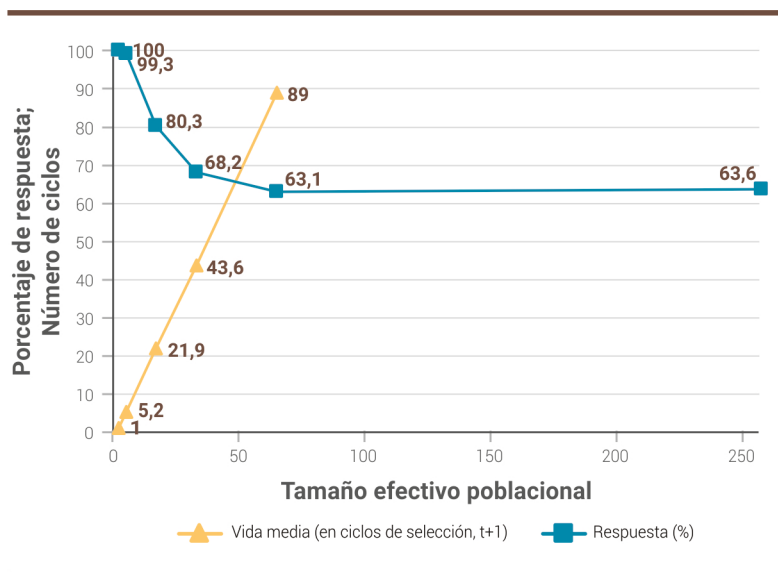


Figura 9.17. Vida media de la selección simulada para poblaciones de diferente tamaño efectivo. El porcentaje de respuesta se refiere al progreso loggable luego de 30 ciclos de selección respecto del límite alcanzable. El número de ciclos se refiere a la vida media de la selección. Adaptado de Baker y Curnow (1969).

Más arriba se indicó cual es el cambio total de frecuencia génica esperable luego de t ciclos de selección:

$$p_t - p_0 = \Delta_p = \sum_{j=0}^t spq [1 - (1/2N)]^j \approx 2Nspq(1 - e^{-t/2N})$$

Para valores medios o grandes del número de genotipos seleccionados (N) la expresión del cambio total de frecuencia génica será:

$$\Delta_p \approx 2Nspq$$

La mitad del cambio de frecuencia génica (vida media de la selección) se alcanzará en $1,4N$ ciclos selectivos (Wricke y Weber, 1986).

Una dificultad muy importante para definir el límite de selección surge de evidencias empíricas. Puede ocurrir que se alcance un límite aparente y que después de alcanzado ese *plateau* se obtengan respuestas positivas a ciclos adicionales de selección. Ello podría explicarse por el surgimiento de nueva variabilidad como consecuencia de mutaciones, o bien de la recombinación entre loci estrechamente ligados en repulsión y mantenidos en un sistema de letales balanceados (Falconer, 1981). En otros casos puede observarse que una vez alcanzado un límite para la selección direccional, se verifique respuesta efectiva a la selección en reversa (Dudley, 2007).

Cuando la selección se practica sobre varios caracteres de manera direccional, la estructura de correlación genética entre ellos y la modalidad de selección simultánea pueden provocar que el límite real de selección para cada carácter individual sea inferior a su respectivo límite teórico.

La disminución progresiva de la variabilidad genética conforme nos acercamos a la fijación de los alelos favorables debería estar acompañada por una disminución de la variancia fenotípica. En algunas situaciones concretas de selección, sin embargo, se observa que la variancia fenotípica, lejos de disminuir, aumenta. De acuerdo con Falconer y Mackay (2001), esta circunstancia puede deberse a diversas causas:

- Efecto de escala causado por la correlación positiva entre variancia fenotípica y media.
- Incremento del componente ambiental de la variancia fenotípica, causado por la mayor sensibilidad ambiental inherente a los individuos homocigotas favorables, cuya frecuencia genotípica se incrementa a medida que la selección resulta efectiva.
- Incremento del componente de variabilidad genética de la variancia fenotípica.

La permanencia o surgimiento de variabilidad genética cuando se supone haber llegado al límite y que hace posible la obtención de respuesta a la selección direccional opuesta, o la recuperación de la respuesta en la misma dirección pueden deberse, entre otras, a las siguientes causas:

- Los alelos desfavorables han disminuido su frecuencia, pero la selección sobre ellos resulta poco efectiva por cuanto estos loci contribuyen en mayor medida a la variancia de dominancia que a la variancia aditiva.
- La selección natural resulta direccional en sentido opuesto a la selección artificial.
- La selección natural y artificial favorecen a los individuos heterocigotas.

Cuando se aplica un método de selección y a los fines prácticos, nos debe interesar estimar la respuesta máxima alcanzable en relación con el desvío estándar aditivo de la población, es decir, la relación:

$$R/\sigma_A \quad [9.15]$$

Si fuese posible conocer con precisión el efecto y número de loci que controlan el carácter bajo selección, podríamos estimar el límite real de selección, pero en vista de que ello no es posible debemos conformarnos solo con pronosticar o predecir el límite teórico a la selección. Para un carácter controlado por n loci, cada uno con efectos del alelo favorable igual a “a” y del alelo desfavorable igual a “-a”, el rango de valores genotípicos posibles en la población será $(-na, na)$ y la respuesta total, medida como la diferencia entre la respuesta a favor y en contra del carácter será entonces (Falconer y Mackay, 2001):

$$R = 2na \quad [9.16]$$

Si la población en selección posee frecuencias génicas intermedias ($p=q=0,5$), su variancia aditiva inicialmente será:

$$\sigma_A^2 = (\frac{1}{2})na^2 \quad [9.17]$$

Por lo tanto elevando al cuadrado la expresión [9.16] nos queda:

$$R^2/\sigma_A^2 = 4n^2a^2/(\frac{1}{2})na^2 = 8n \quad [9.18]$$

En conclusión, la respuesta será mayor cuanto mayor sea el número de loci segregantes. Sin embargo, si el tamaño efectivo poblacional no es lo suficientemente grande como para aminorar la pérdida aleatoria de alelos favorables antes de que tengan oportunidad de ser seleccionados, la respuesta máxima posible no podrá alcanzarse. Recordemos que la probabilidad de perder el alelo favorable a causa de la deriva genética es inversamente proporcional a su frecuencia génica inicial; en otras palabras, cuanto más raro sea el alelo, mayor será la probabilidad de su pérdida por causas aleatorias. Pero la probabilidad de fijación del alelo cuando se ejerce selección a su favor, será mayor cuanto mayor sea el producto entre el coeficiente de selección (s) y el tamaño efectivo poblacional. En consecuencia, la mejor estrategia para alcanzar el límite máximo por selección será utilizar esquemas selectivos con alta intensidad de selección aplicadas en poblaciones con gran tamaño efectivo (N_e). Robertson (1960) citado por Falconer indica que la respuesta máxima alcanzable por selección (${}^L R$) para un carácter bajo control exclusivamente aditivo ($d=0$) es:

$${}^L R = 2N_e R = 2N_e i h^2 \sigma_p \quad [9.18]$$

El límite por selección será aún mayor para el caso de $d < 0$ (alelos favorables son parcialmente recesivos). Esta expresión tiene únicamente valor teórico porque siempre existirán limitantes de orden práctico respecto del tamaño efectivo poblacional y será imposible evitar la pérdida de alelos favorables.

Para estudiar el límite a la selección artificial, además de la consideración sobre la respuesta máxima alcanzable, hay otros abordajes, como es tener en cuenta el comportamiento esperado del mejor genotipo obtenible a partir de un determinado esquema de selección, la probabilidad de fijación del alelo favorable y la proporción de loci que lograron fijar el alelo favorable cuando todos ellos alcanzaron la homocigosis.

Bailey (1977) comparó los mejores genotipos extraíbles de una población sometida a dos estrategias de selección: selección recurrente y selección no recurrente. En el caso de la selección no recurrente (se volvía a muestrear la misma población en su condición original en cada ciclo) observó que luego de cuatro a cinco ciclos de selección no recurrente la respuesta alcanzaba un *plateau*. En el caso de selección recurrente hay evidencias empíricas que destacan la persistencia de la respuesta a la selección luego de un número considerable de ciclos selectivos (Dudley, 2007).

Cuando se desarrollan líneas puras de una especie autógama o líneas endocriadas de especies alógamas a partir de una población de cría biparental ($p=q=0,5$) es muy importante considerar el límite alcanzable por selección en términos del mejor genotipo resultante. El objetivo es que al término de la selección se haya logrado combinar en los nuevos genotipos (nuevas líneas) la mayor cantidad posible de los alelos favorables provenientes de uno y de otro progenitor de la población biparental. Si bien todos los alelos favorables estarán presentes en la F_1 , el objetivo es obtener individuos homocigotas que posean el máximo número de alelos favorables en todos los loci y, por supuesto, que superen a cualquiera de los progenitores de la población biparental. Supongamos que de la población F_2 se derivaran líneas aplicando el método de conducción de una única semilla (*Single Seed Descent* o SSD) y pretendemos conocer la probabilidad de lograr líneas que posean el alelo favorable en condición homocigota en todos los loci. De acuerdo a Bailey (1977), la probabilidad de que al menos m loci hayan fijado el alelo favorable sobre un total de $n=30$ loci segregantes en la población F_2 y en ausencia de ligamiento, puede calcularse a partir de la siguiente expresión:

$$P_{(m)} = (n! / m!(n-m)!) P^m (1-P)^{n-m} \quad [9.19]$$

donde P es la probabilidad de que un locus contenga el alelo favorable en la línea homocigota, que a su vez resulta ser:

$$P = f_1 P_1 + f_2 P_2,$$

donde:

P_1 es la probabilidad de que el alelo favorable resulte fijado en la línea homocigota, habiendo estado ese locus en estado heterocigota en la filial F_1 .

f_1 es la probabilidad de que las plantas F_1 sean heterocigotas para el locus considerado.

P_2 es la probabilidad de que el alelo favorable resulte fijado en la línea homocigota, habiendo estado ese locus en estado homocigota para el alelo favorable en la filial F_1 .

f_2 es la probabilidad de que las plantas F_1 sean homocigotas para el alelo favorable para el locus considerado.

Aquellas probabilidades $P(m)$ se presentan en la Tabla 9.6.

Número total de loci (M)	P(m)	P(m o >m)
20	0,115	0,209
22	0,051	0,094
24	0,012	0,017
25	0,004	0,006
26	0,001	0,002

Tabla 9.6. Probabilidad de ocurrencia en un mismo genotipo de loci homocigotas (M) para el alelo favorable (m o mayor que m) cuando el número total de loci es 30.

De acuerdo a la tabla anterior, la probabilidad de recuperar por selección una línea que combine en sus 30 loci los alelos favorables provenientes de uno y otro padre es prácticamente cero. La probabilidad de seleccionar individuos que tengan al menos 20 loci fijados para los alelos favorables es aproximadamente el 21 %. La probabilidad de obtener una línea que sea superior al mejor progenitor también dependerá de la distribución de los alelos favorables entre los padres. Cuando los padres muestran diferencias importantes en la distribución de alelos puede convenir seleccionar sobre la generación de retrocruza al mejor progenitor más que sobre la filial F_2 .

9.12. Número de factores efectivos y respuesta a la selección

En la medida que la respuesta a la selección para un carácter depende del número de loci que lo controla, que estos tengan efecto similar y segreguen independientemente unos de otros, sería teóricamente posible inferir su número a partir de la respuesta a un experimento de selección. Pero difícilmente se verifique en la realidad el supuesto de efectos similares a través de loci y por otra parte el número de loci cuantitativos involucrados hace muy probable que al menos una fracción de ellos esté físicamente ligada en el mismo cromosoma. En tal circunstancia la selección favorecerá segmentos cromosómicos (haplotipos) que contienen varios loci ocupados por alelos favorables. A cada uno de esos segmentos que segregan en forma conjunta y que están expuestos a la selección se lo denomina factor efectivo. El número de factores efectivos (e_n), podría estimarse a partir de la respuesta a la selección siguiendo la expresión (Falconer y Mackay, 2001):

$$R^2/\sigma_A^2 = (8)(e_n),$$

por lo tanto

$$e_n = R^2/8\sigma_A^2$$

En este caso R es la respuesta total a la selección divergente, es decir, la diferencia entre la respuesta a favor y en contra del carácter. Se asume que las frecuencias génicas en la población de partida son $p=q=0,5$ y que los efectos de los factores efectivos son iguales entre sí.

9.13. Estimación de respuesta a la selección

Existe una amplia serie de métodos de estimación de la respuesta a la selección, cuya complejidad deriva de cuales son las respuestas y efectos que se deseen estimar. Uno de los modelos más sencillos es el de Eberhart (1964) que aplica un método de mínimos cuadrados y estima las respuestas a la selección como una regresión respecto de los ciclos de selección. Otros modelos más sofisticados se aplican especialmente para investigar los efectos de la selección recurrente sobre la deriva genética, la contribución de los efectos de aditividad y no aditividad a la respuesta observada (Smith, 1979a, 1979b, 1983, 1984; Eyherabide y Hallauer, 1991b), las modificaciones en la habilidad combinatoria y heterosis (Hammond y Gardner, 1974) y la divergencia entre las poblaciones (Hanson, 1983, 1987). Hanson y Moll (1986) compararon las modificaciones que ocurren en la contribución de efectos de dominancia en poblaciones bajo selección interpoblacional e intrapoblacional.

9.14. Identificación y selección de progenitores de poblaciones de cría

Diversos autores han propuesto metodologías para asistir la elección de líneas o poblaciones progenitoras de poblaciones de cría (Dudley, 1984a, 1984b, 1987; Gerloff y Smith, 1988; Zanoni y Dudley, 1989; Bernardo, 1990. Describiremos a continuación la metodología propuesta por Dudley (1984, 1987).

9.14. 1. Modelos de Dudley

La identificación de germoplasma útil para incorporar en programas de mejoramiento es siempre compleja, pero lo es más cuando el objetivo es mejorar caracteres de herencia cuantitativa. A diferencia de la identificación de fuentes de variabilidad genética para mejorar características de herencia relativamente simple (como por ejemplo algunos tipos de resistencias a factores adversos, o atributos de calidad), en los de herencia cuantitativa la menor correspondencia entre la expresión fenotípica y la constitución genotípica es un inconveniente. Esa dificultad puede aliviarse mediante la genotipificación (en caso de disponerse de marcadores previamente identificados y de gran efecto) o mediante la aplicación de modelos cuantitativos que permitan identificar el germoplasma que contenga alelos favorables que no se encuentran presentes en las colecciones de trabajo de los programas de mejoramiento.

Dudley presentó sendos modelos para identificar líneas endocriadas y poblaciones segregantes que poseen "nuevos" alelos que permitirían desarrollar nuevas líneas endocriadas (a partir de cruzamientos línea x línea o población x línea) que resulten en la superación de un híbrido considerado élite o de referencia en determinado rango ambiental.

El modelo para identificar líneas endocriadas (Dudley, 1984a) que poseen nuevos alelos asume dominancia completa ($d=a$) y que a (valor aditivo respecto al promedio de los dos progenitores homocigotas) es constante a través de loci. Dada la existencia de un híbrido de referencia $L_1 \times L_2$ y que existe una serie de líneas alternativas, denominadas genéricamente L_z , pueden describirse ocho clases de loci (A, B, C, D, E, F, G y H) según sean las condiciones de ausencia/presencia de alelos entre las tres líneas involucradas (Tabla 9.7).

Clases de loci	Líneas		
	L_1	L_2	L_z
A	+	+	+
B	+	+	-
C	+	-	+
D	+	-	-
E	-	+	+
F	-	+	-
G	-	-	+
H	-	-	-

Tabla 9.7.

Clasificación de loci homocigotas para alelos favorables o desfavorables en función de su presencia en las dos líneas parentales de un híbrido élite o de referencia (L_1 x L_2) y de terceras líneas (L_z) que podrían servir de fuente de nuevos alelos favorables no presentes en las primeras.

El modelo requiere evaluar una serie de medias generacionales, a saber: L_1 , L_2 , L_z , $L_1 \times L_2$, $L_1 \times L_z$, $L_2 \times L_z$, y asumir que el número de alelos en la clase A es igual al de la clase H (Tabla 9.8).

Generación	Clases de loci						Valor	
	B	C	D	E	F	G	Genotípico	Fenotípico
L_1	1	1	1	-1	-1	-1	$a(B+C+D-E-F-G)$	X_{11}
L_2	1	-1	-1	1	1	-1	$a(B+E+F-C-D-G)$	X_{12}
L_z	-1	1	-1	1	-1	1	$a(C+E+G-B-D-F)$	X_{1z}
$L_1 \times L_2$	1	1	1	1	1	-1	$a(B+C+D+E+F-G)$	$X_{11 \times 12}$
$L_1 \times L_z$	1	1	1	1	-1	1	$a(B+C+D+E+G-F)$	$X_{11 \times 1z}$
$L_2 \times L_z$	1	1	-1	1	1	1	$a(B+C+E+F+G-D)$	$X_{12 \times 1z}$

Tabla 9.8.

Valores fenotípicos y genotípicos esperados según el modelo de Dudley para líneas y cruzamientos.

Se establece entonces un sistema de ecuaciones (una por cada media generacional) con seis incógnitas, donde el número de loci que componen la

clase G es la de mayor interés puesto que corresponde a alelos presentes en la línea L_z , que no están presentes en el híbrido de referencia $L_1 \times L_2$.

De manera análoga al análisis de medias generacionales vistos en capítulos anteriores, los parámetros del modelo de Dudley pueden estimarse a partir de un sistema de ecuaciones donde cada uno de ellos se presenta como una combinación lineal de las medias de los valores fenotípicos de las generaciones (Tabla 9.9).

Parámetro	Soluciones al sistema
aB	$[(Y_{11} \times Y_{12}) + (Y_{12} \times Y_{12}) + Y_{11} + Y_{12} - Y_{12} - (Y_{11} \times Y_{12})]/4$
aC	$[(Y_{11} \times Y_{12}) + (Y_{12} \times Y_{12}) + Y_{11} + Y_{12} - Y_{12} - (Y_{11} \times Y_{12})]/4$
aD	$[(Y_{11} \times Y_{12}) + (Y_{11} \times Y_{12}) - Y_{11} - Y_{12} - Y_{12} - (Y_{12} \times Y_{12})]/4$
aE	$[(Y_{11} \times Y_{12}) + (Y_{11} \times Y_{12}) - Y_{11} + Y_{12} + Y_{12} - (Y_{12} \times Y_{12})]/4$
aF	$[(Y_{12} \times Y_{12}) + (Y_{11} \times Y_{12}) - Y_{11} - Y_{12} - Y_{12} - (Y_{11} \times Y_{12})]/4$
aG	$[(Y_{11} \times Y_{12}) + (Y_{12} \times Y_{12}) - Y_{11} - Y_{12} - Y_{12} - (Y_{11} \times Y_{12})]/4$

Tabla 9.9. Soluciones para los grupos de loci a partir de los valores fenotípicos de los cruzamientos.

Dado un número de líneas alternativas (L_z) convendrá utilizar aquella que tiene el grupo de loci tipo G más numeroso. Una estimación positiva de aG significa que L_z puede aportar alelos no presentes en ninguna de las líneas progenitoras del híbrido de referencia. Estimaciones positivas de aD indican que L_1 posee alelos favorables que están ausentes en L_2 y L_z , mientras que estimaciones positivas de aF indican que L_2 posee loci con alelos favorables ausentes en L_1 y L_z . Los restantes parámetros sirven para indicar las coincidencias a nivel genotípico de L_z con L_1 y con L_2 . La suma de aC + aF o de aD + aE indican, respectivamente, la similitud a nivel de loci entre L_1 y L_z y entre L_2 y L_z .

La obtención de estimaciones del número de loci menores que cero puede ocurrir por simples deficiencias experimentales, pero si resultaran significativamente menores que cero, ello podría evidenciar que el supuesto de dominancia completa ($d=a$) no se verifica. En tal caso, si $d \neq a$, las esperanzas de las estimaciones de los parámetros serían diferentes, conforme se presenta en la Tabla 9.10.

Parámetro	Esperanza de las estimaciones si $d \neq a$
aB'	$[B(a+ad) - G(a -ad)]/2$
aC'	$[C(a+ad) - F(a -ad)]/2$
aD'	$[D(a+ad) - E(a -ad)]/2$
aE'	$[E(a+ad) - D(a -ad)]/2$
aF'	$[F(a+ad) - C(a -ad)]/2$
aG'	$[G(a+ad) - B(a -ad)]/2$

Tabla 9.10. Parámetros y esperanzas de las estimaciones del modelo de Dudley en el caso de que no exista dominancia completa.

En caso de que el supuesto de que $A=H$ no se cumpla, la esperanza de las estimaciones si $A \neq H$ pasan a ser las que se presentan en la Tabla 9.11.

Parámetro	Esperanza de las estimaciones si $A \neq H$
aB''	$aB + (aA-aH)/2$
aC''	$aC + (aA-aH)/2$
aD''	$aD - (aA-aH)/2$
aE''	$aE - (aA-aH)/2$
aF''	$aF - (aA-aH)/2$
aG''	$aG - (aA-aH)/2$
$aC'' + aF''$	$aC + aF$
$aD'' + aE''$	$aD + aE$

Tabla 9.11. Parámetros y esperanzas de las estimaciones del modelo de Dudley en el caso de que $A \neq H$.

El modelo para identificar poblaciones portadoras de alelos favorables no presentes en ninguna de las líneas progenitoras de un híbrido de referencia es diferente al anterior (Dudley, 1984b; 1987). En primer lugar existirán apenas 4 clases de loci, de acuerdo a la Tabla 9.12.

Clases de loci	Líneas parentales del híbrido		Población
	L_1	L_2	P_z
I = (A+B)	1	1	P
J = (C+D)	1	0	P
K = (E+F)	0	1	P
L = (G+H)	0	0	p_L

Tabla 9.12.

Clasificación de loci homocigotas para alelos favorables o desfavorables en función de su presencia en las dos líneas parentales de un híbrido élite o de referencia ($L_1 \times L_2$) y de poblaciones (P_z) que podrían servir de fuente de nuevos alelos favorables no presentes en las primeras.

Se asume que las poblaciones alternativas indicadas como P_z están en equilibrio de Hardy-Weinberg. Las esperanzas de las estimaciones de las medias de las generaciones L_1 , L_2 , P_z , $L_1 \times P_z$, $L_2 \times P_z$ e $L_1 \times L_2$ serán las que se indican en Tabla 9.13.

Genotipos	Valor genotípico
L_1	$a(I+J-K-L)$
L_2	$a(I-J+K-L)$
P_z	$a(I+J+K)(4p-2p^2-1) + aL(4p_L-2p_L^2-1)$
$L_1 \times P_z$	$a(I+J + 2Kp -K + 2Lp_L -L)$
$L_2 \times P_z$	$a(I -J + 2Jp +K + 2Lp_L -L)$
$L_1 \times L_2$	$a(I+J+K-L)$

Tabla 9.13.

Valores fenotípicos y genotípicos esperados según el modelo de Dudley para líneas, poblaciones alternativas y cruzamientos entre líneas y poblaciones.

El parámetro que más interesa estimar es el p_L correspondiente a cada una de varias poblaciones candidatas, a fin de identificar aquella que posea mayor frecuencia de alelos favorables no presentes en las dos líneas progenitoras. Las soluciones al sistema de ecuaciones se presentan en la Tabla 9.14.

Parámetro	Estimador
Z	$[(l_1 \times P_z) - l_1][l_1 \times l_2 - l_2] - [(l_2 \times P_z) - l_2][l_1 \times l_2 - l_1]$
b	$2(l_1 - l_2)[l_1 \times l_2(1 - 2q^2) - P_z] + 4Z$
aJ	$[(l_1 \times l_2) - l_2]/2$
aK	$[(l_1 \times l_2) - l_1]/2$
q=1-p	$[(l_1 \times P_z) - (l_2 \times P_z)] / (l_1 - l_2)$
p_L	$[-b - (b^2 - 16q^2Z^2)^{0.5}] / -4Z$
aL	$-4Z^2 / 2(l_1 - l_2) [-b - (b^2 - 16q^2Z^2)^{0.5}]$
al	$aL + (l_1 + l_2) / 2$
aLp_L	$Z / 2(l_1 - l_2)$

Tabla 9.14. Soluciones de los grupos de loci a partir de los valores fenotípicos de los cruzamientos.

El parámetro de mayor interés resulta ser aLp_L , cuya interpretación no requiere el supuesto de equilibrio poblacional.

El modelo de Dudley permite obtener estimadores de parámetros orientadores sobre la mejor estrategia para incorporar los nuevos alelos que se identifiquen en L_z o P_z . El proceso se inicia con el cruzamiento de alguna de las dos líneas de referencia (L_1, L_2) con el material donante (L_z o P_z) y posteriormente por la autofecundación de la F_1 o por retrocruzamiento a la línea elegida como recurrente.

Dudley (1982) desarrolló la teoría para estos casos. Considerando dos alelos por locus, ausencia de ligamiento y de epistasis, partiendo de una población resultante del cruzamiento entre un progenitor donante (D) y otro receptor (o recurrente) (R), podemos considerar dos clases de loci. La primera clase de loci (r) corresponde a aquellos para los cuales R es homo-

cigota para el alelo favorable (AA), mientras que D es Aa o aa. La segunda clase de loci (d) está integrada por aquellos donde R es aa y D es AA o Aa.

Los efectos de cada genotipo serán:

AA: a

Aa: d

aa: -a

La media de la población en apareamiento aleatorio será:

$$\mu = D[a(p_d - q_d) + 2dp_dq_d] + R[a(p_r - q_r) + 2dp_rq_r],$$

donde D y R es el número de loci de las clases D y R, respectivamente, y p_r , q_r , p_d y q_d son las frecuencias de los alelos desfavorables y favorables dentro de las clases de loci R y D, respectivamente.

La media de la población de cría y la de las líneas derivables de ella por endocría y selección dependerán de la proporción relativa de las dos clases de loci (D y R) y de las frecuencias génicas en cada una de ellas. Puede esperarse que cuanto mayor sea la relación entre el número de loci del tipo D respecto del tipo R y mayor sea la frecuencia de los alelos favorables en el grupo D, mayores serán la media poblacional y la de las líneas derivadas de esta.

La media de las líneas derivables de esa población en ausencia de selección y la variancia entre ellas serán:

$$\mu_L = Da(p_d - q_d) + Ra(p_r - q_r)$$

$$\sigma_L^2 = D4a^2p_dq_d + R4a^2p_rq_r$$

Si el donante D y el receptor R son ambos líneas homocigotas, la nueva línea puede derivarse desde la F_2 del cruzamiento $D \times R$, $[(D \times R)F_2]$ directamente por endocría y selección, o bien practicarla sobre la generación resultante de una o más generaciones de retrocruza al padre recurrente, $[(D \times R)R^n]$.

La variancia aditiva en la población será:

$$\begin{aligned} \sigma_A^2 &= D\sigma_{Ad}^2 + R\sigma_{Ar}^2 = D2p_dq_d\alpha_d^2 + R2p_rq_r\alpha_r^2 = \\ &= D2p_dq_d[a + (q_d - p_d)d]^2 + R2p_rq_r[a + (q_r - p_r)d]^2 \end{aligned}$$

La respuesta a la selección vendrá dada por el balance entre el incremento de la frecuencia de los alelos favorables a expensas de los desfavorables en ambas clases de loci, es decir, que la selección será en el sentido buscado cuanto mayor sea la relación entre el aumento de la frecuencia de los alelos pertenecientes a la clase de loci D, (Δ_d), respecto del aumento de la frecuencia de los alelos pertenecientes a la clase R, (Δ_r). De acuerdo a Dudley (1982), los cambios de frecuencias serán:

$$\Delta_d = ip_d q_d \alpha_d / \sigma_p$$

$$\Delta_r = ip_r q_r \alpha_r / \sigma_p$$

$$\Delta_d / \Delta_r = [1 + d(q_d - p_d)] / [1 + d(q_r - p_r)]$$


La estrategia de selección más efectiva será iniciarla en la generación en la que se espera que la relación entre el cambio de frecuencias génicas sea la mayor posible.

Excepto para el caso en que $d=0$, la relación entre ambas repuestas será mayor a 1 en cualquier generación de retrocruzamiento y creciente hasta la RC_3 . Paralelamente aumentará p_r en la medida que se hacen más retrocruzamientos (Tabla 9.15).

Generación	Frecuencia de alelo favorable		Nivel de dominancia		
	p_r	p_d	0,0	0,5	1,0
F_2	0,5	0,5	1	1	1
$RC_1 F_2$	0,75	0,25	1	1,67	3
$RC_2 F_2$	0,875	0,125	1	2,20	7
$RC_3 F_2$	0,9375	0,0625	1	2,56	15

Tabla 9.15. Generación a partir de la cual se inicia la selección buscando genotipos que reúnen la mayor proporción posible de alelos favorables del donante, para tres niveles de dominancia (aditividad completa, dominancia parcial y dominancia completa. Basado en Dudley (1982).

Siempre que el número de loci de la clase R sea mayor que el número de loci de la clase D, se espera que la media de la población aumente con el número de generaciones de retrocruzamiento. Por lo tanto convendría rea-



lizar al menos una generación de retrocruzamiento previa a la endocría y selección. Toda vez que incrementa la frecuencia p_r , p_d irá disminuyendo. Las variancias aditivas máximas ocurren cuando $p=q=0,5$ si $d=0$ y cuando $p=0,25$ si $d=a$. Por lo tanto la variancia aditiva para los loci de tipo D serán menores que el máximo para RC_2F_2 o RC_3F_2 . De cualquier manera, debe tenerse en consideración el tamaño efectivo poblacional. Si el tamaño efectivo es reducido, hacer más de una retrocruza no resultaría aconsejable (Ho y Comstock, 1980).

Empleando un modelo de simulación Dudley (1982) analizó en cuál generación de retrocruzamiento convendría iniciar la selección para mejorar la población biparental y desarrollar mejores líneas endocriadas a partir de ella. Asimismo investigó sobre el efecto de la selección previa a iniciar la endocría, ya sea sobre la F_2 o sobre generaciones de retrocruzamiento, cuando el objetivo es obtener líneas que superen a la mejor de las líneas progenitoras de la población de cría original. Las variables que tuvo en cuenta fueron la naturaleza de las generaciones (F_2 , RC_1F_2 , RC_2F_2 , RC_3F_2), diferentes niveles de dominancia ($d=0$, $d=0,5$ y $d=1$) y niveles de divergencia variables entre los dos progenitores cuantificado por sus proporciones relativas de número de loci homocigotas para los alelos favorables (70-30 hasta 95-5) (Tabla 9.16).

Cuando el fitomejorador está interesado en mejorar la media de una población (Tabla 9.16 [A]), será conveniente iniciar la selección sobre alguna generación de retrocruzamiento. Convendrán más generaciones de retrocruzamiento cuanto más divergentes sean los progenitores y cuanto mayor sea el nivel de dominancia. En el caso de dominancia completa y menor divergencia entre los padres (70-30) resultaría indistinto iniciar la selección en la F_2 o en la generación de retrocruza. Sin embargo, como para generar cada retrocruzamiento hay que asignar tiempo y otros recursos resultaría más conveniente trabajar sobre la F_2 .

En cuanto a la probabilidad de que al menos el 25 % de las líneas derivables al azar de cada tipo de población superen al mejor progenitor, el número de ciclos de selección previa tiende a ser menor mientras más generaciones de retrocruzamiento se realicen, menos divergentes sean los progenitores y menor sea el desvío de dominancia.

[A] Media poblacional												
DIV	F₂			RC₁F₂			RC₂F₂			RC₃F₂		
	a=0	a=0,5	a=1	a=0	a=0,5	a=1	a=0	a=0,5	a=1	a=0	a=0,5	a=1
70-30	5	2	0	3	1	0	3	0	0	2	0	0
80-20	8	5	2	5	3	0	4	2	0	4	1	0
95-5	15	14	6	10	9	4	8	7	1	7	5	0

[B] Líneas homocigotas												
DIV	F₂			RC₁F₂			RC₂F₂			RC₃F₂		
	a=0	a=0,5	a=1	a=0	a=0,5	a=1	a=0	a=0,5	a=1	a=0	a=0,5	a=1
70-30	4	5	5	2	3	3	2	2	2	1	1	1
80-20	7	8	10	4	5	6	3	3	4	2	2	3
95-5	13	17	>25	9	12	25	6	8	20	5	6	8

Tabla 9.16.


[A] Número de ciclos de selección requeridos para que la media de la población seleccionada sea mayor o igual a la media del mejor progenitor para diferentes desvíos de dominancia y divergencia (DIV) entre los padres. [B] Número de ciclos de selección requeridos para que el 25 % de las líneas derivadas de la población tengan medias superiores a la media del mejor progenitor, para distintos niveles de dominancia y divergencia (DIV) entre las líneas progenitoras. Basado en Dudley, 1982.

Puede estimarse la frecuencia de líneas homocigotas extraíbles del ciclo 0 de la población original que tendrían mejor comportamiento que la mejor línea progenitora (Tabla 9.17).

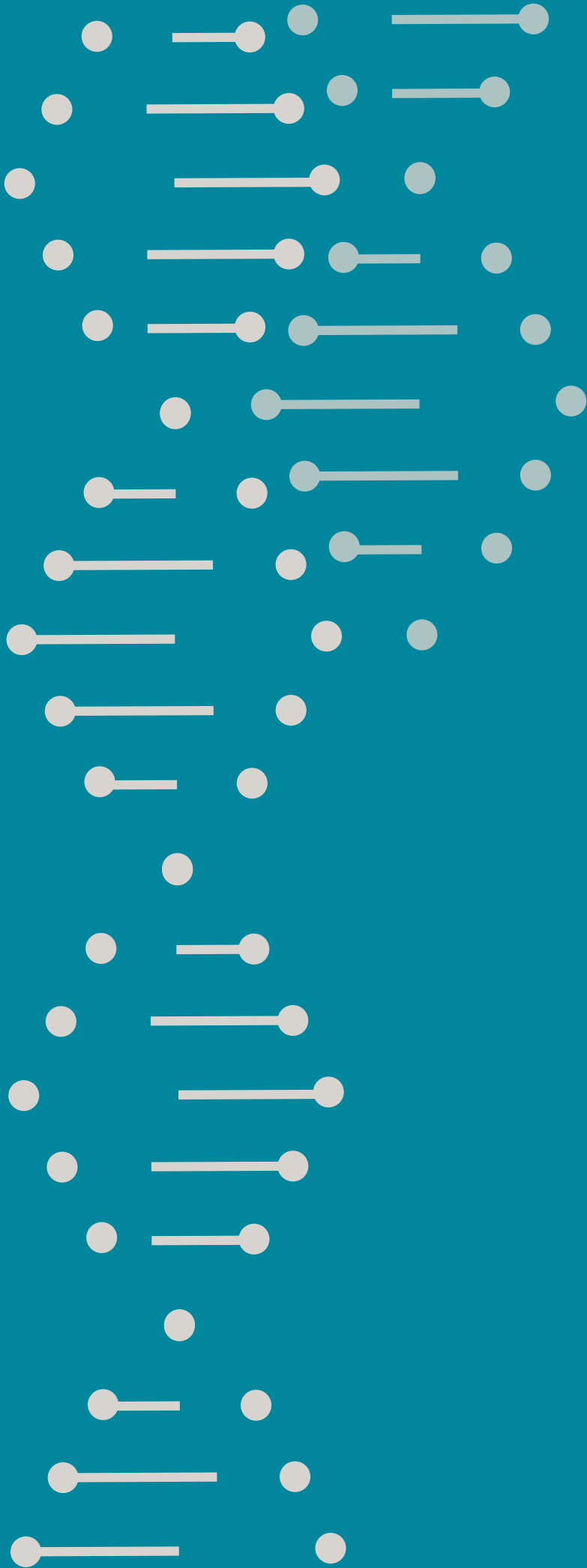
DIV	Generaciones			
Clases R – D	F₂	RC₁F₂	RC₂F₂	RC₃F₂
70-30	<0,1	10	67	147
80-20	<0,1	0,3	12	57
95-5	<0,1	<0,1	0,3	9

Tabla 9.17.

Frecuencia de líneas homocigotas (por cada 1000) extraídas al azar de poblaciones F₂ y retrocruzas derivadas de un cruzamiento biparental entre líneas homocigotas con diferente grado de divergencia entre ellas (DIV), que serán superiores a la mejor de las líneas progenitoras. Basado en Dudley, 1982.



Puede concluirse que hacer más generaciones de retrocruzamiento aumentará la probabilidad de generar nuevas líneas superiores a los dos progenitores, especialmente cuando más genéticamente divergentes sean entre sí.



Selección recurrente

10.1. Objetivo de la selección recurrente

La selección recurrente de una población consiste en la evaluación sistemática de individuos o sus familias, la recombinación de las selecciones para formar la población mejorada y la repetición del proceso con el objetivo de aumentar la media poblacional para el carácter de interés con mínima afectación de la variancia genética (Figura 10.1). El aumento de la media poblacional responderá a incrementos en la frecuencia de los alelos favorables que controlan el carácter.

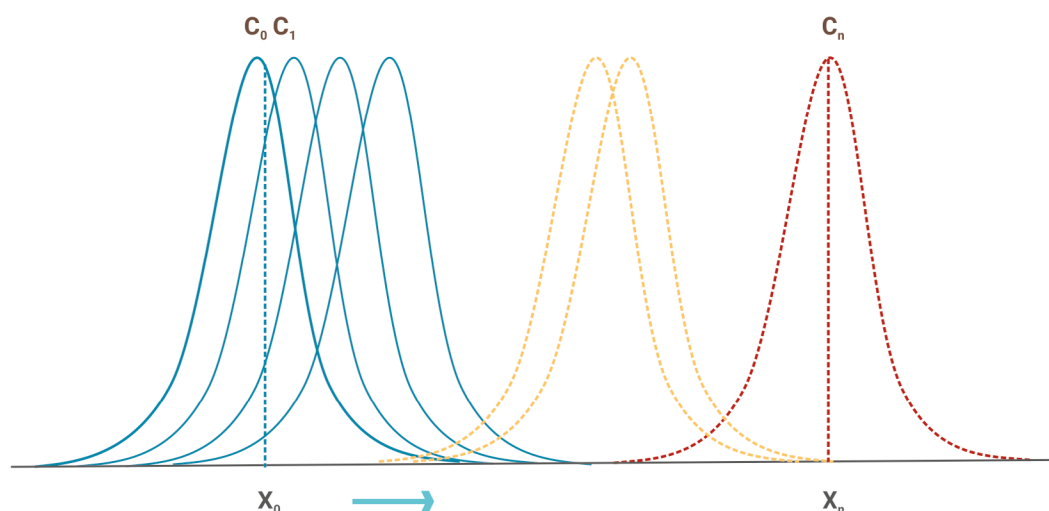


Figura 10.1.

Representación de la respuesta ideal a n ciclos de selección recurrente para un carácter cuantitativo bajo selección direccional en una población original (C_0), en términos de su distribución de frecuencia en ciclos sucesivos ($C_1 \dots C_n$). La flecha indica el sentido de la selección.

En el Capítulo 2 señalamos que cuanto mayor es el número de loci que controlan el carácter, mayor es el número de genotipos y fenotipos posibles (Tabla 10.1).

N.º de loci independientes (n)	N.º de genotipos posibles o de fenotipos posibles con dominancia parcial (3^n)	N.º de fenotipos posibles con dominancia completa (2^n)
2	9	4
5	19,683	64
10	387.420.489	4,096
20	$1,5 \times 10^{17}$	16.777.216

Tabla 10.1.

Número de genotipos y fenotipos posibles de encontrar en una población de tamaño infinito para diferente número de loci (n) que controlan un carácter cuantitativo, bajo un modelo genético de dominancia parcial y otro de dominancia completa.

El aumento ocurre de manera exponencial. Esto plantea al fitomejorador dos dificultades prácticas: en primer lugar, cómo disponer de una muestra de individuos del tamaño suficiente como para garantizar, con cierto nivel de probabilidad, que al menos una planta con el genotipo deseado o buscado estará incluida en dicha muestra. En segundo lugar, asumiendo que el tamaño de muestra fuera el suficiente, cómo lograr la identificación de ese genotipo dentro del conjunto de individuos y consecuentemente seleccionarlo.

En una población en equilibrio de Hardy-Weinberg y para un carácter controlado por n loci ($i=1 \dots n$), asumiendo que las frecuencias génicas (p_i) sean las mismas en cada uno de ellos, dos alelos posibles por locus y que el genotipo ideal sea el que posee los alelos favorables al estado homocigota en los n loci, la frecuencia esperada de ese genotipo ideal (I) en la población será:

$$I = (p_i^2)^n$$

La probabilidad de que un individuo tomado al azar de esa población no posea el genotipo ideal será: $1 - I$.

Para un determinado número de loci, a medida que aumenta la frecuencia del alelo favorable, se incrementará la frecuencia del genotipo ideal en la población (Tabla 10.2). Nótese que el efecto del incremento de la frecuencia génica sobre la probabilidad o frecuencia del genotipo ideal es mayor cuanto mayor es el número de loci que controla el carácter.

N.º de loci	Frecuencia del alelo favorable (p)						
	0,60	0,70	0,75	0,80	0,85	0,90	0,95
n							
10	0,000037	0,0008	0,0037	0,01153			
14		0,000046	0,00032	0,00193			
16		0,000011	0,00010	0,00079	0,0055		
20			0,00001	0,00013	0,0015		
40					0,0000023	0,00022	0,0165

Tabla 10.2. Probabilidad del genotipo ideal (I) para distintas frecuencias del alelo favorable (p) y diferente número de loci (n) segregantes.

Como se señaló más arriba, a los fines prácticos de un programa de mejoramiento, al fitomejorador le interesa poder incluir a los genotipos ideales dentro de muestras poblacionales de tamaños que resulten operativamente manejables para luego identificarlos y seleccionarlos. Obsérvese en la Tabla 10.3 que conforme aumenta la probabilidad del genotipo ideal, aumenta notablemente la probabilidad de haber incluido al menos un individuo con el genotipo buscado en muestras de tamaños relativamente manejables a nivel práctico.

Probabilidad del genotipo ideal (I)	Tamaño de muestra (número de individuos, m)			
	50	100	200	400
0,00001	0,0005	0,001	0,0020	0,0040
0,00005	0,0025	0,005	0,0100	0,0198
0,0001	0,005	0,010	0,0198	0,0392
0,0005	0,025	0,049	0,095	0,181
0,001	0,049	0,095	0,181	0,330
0,003	0,139	0,260	0,452	0,699
0,005	0,220	0,390	0,630	0,87
0,1	0,390	0,630	0,870	0,98

Tabla 10.3.

Probabilidad de haber incluido en muestras de tamaño m a por lo menos un individuo con el genotipo ideal cuando su frecuencia en la población total es (I) .

Combinando la información de la Tabla 10.2 con la de la Tabla 10.3 concluimos que pequeños aumentos en la frecuencia de los alelos favorables generan incrementos muy importantes en la probabilidad de encontrar individuos con el genotipo ideal en muestras del tamaño de las que habitualmente se manejan en los programas de mejoramiento vegetal (Comstock, 1996).

La lógica detrás de la selección recurrente es lograr paulatinos incrementos en la frecuencia de los alelos favorables, que tienen como consecuencia el aumento de la media de la población, y el incremento de la probabilidad de encontrar y seleccionar genotipos superiores. Además, a medida que se mejora la población, también aumenta la aptitud de la población como fuente, por ejemplo, de líneas progenitoras de híbridos superiores.

La selección recurrente consta de varias etapas que se repiten en cada ciclo (Figura 10.2):

- i. Muestreo de individuos en la población y formación de familias (progenies).
- ii. Evaluación de individuos (o bien pruebas de progenie a partir de sus familias) y selección.
- iii. Recombinación de los individuos o familias seleccionadas.



Figura 10.2.

Esquema general de las etapas de la selección recurrente a partir de una población original.

La evaluación de los individuos de la población tiene como objetivo disponer de una estimación de sus valores de mejora o reproductivos. Para ello el fitomejorador tiene diversas opciones:

- i. Evaluación fenotípica (*per se*) de plantas individuales.
- ii. Evaluación del valor de mejora de plantas individuales mediante pruebas de progenie. Conceptualmente se trata de una evaluación genotípica de los individuos a partir de la recolección de datos fenotípicos de sus respectivas progenies.
- iii. Evaluación de individuos a partir de datos genotípicos a nivel molecular.
- iv. Evaluación de individuos a partir de índices que integran información fenotípica y genotípica (a nivel de modelos de genética cuantitativa y molecular).

10.2. Clasificación de los métodos de selección recurrente

De acuerdo a la finalidad buscada por la selección recurrente, los métodos se dividen en dos grandes grupos:

- i. Selección recurrente ***intrapoblacional***, cuando el objetivo es mejorar la población para uno o más caracteres.
- ii. Selección recurrente ***interpoblacional***, cuyo objetivo es mejorar el comportamiento del cruzamiento entre dos poblaciones, para uno o más caracteres.

La experiencia indica que los métodos de selección recurrente intrapoblacional, en mayor o menor medida, han sido efectivos en mejorar la media poblacional (respuesta directa). La selección recurrente recíproca ha mostrado ser eficiente en mejorar el cruzamiento interpoblacional (respuesta directa), pero no necesariamente la media de las poblaciones *per se* (respuesta indirecta). Desde el punto de vista del mejoramiento sería deseable que ambos tipos de respuesta fueran positivos ya que como se indicó más arriba, el mejoramiento de las poblaciones *per se* quedará más tarde reflejado en la calidad de las líneas endocriadas que podrán derivarse de ellas (Souza Jr., 1993).

Por la clase de información utilizada en la evaluación de los individuos o familias, podemos clasificar los métodos en:

- i. Selección recurrente fenotípica (o selección masal con recombinación, generalmente en especies alógamas) (Fehr, 1987).
- ii. Selección recurrente (a partir de datos fenotípicos de pruebas de progenie).
- iii. Selección recurrente asistida por marcadores moleculares.
- iv. Selección recurrente genómica.

Los métodos de selección recurrente intrapoblacionales que evalúan familias se clasifican en:

- i. Selección recurrente de familias de medios hermanos (cruzamientos de prueba).

Utilizando la propia población como probadora:

- Selección espiga x hilera.
- Selección espiga x hilera modificada.
- Selección espiga x hilera doble modificada.

Utilizando un probador externo o no relacionado con la población:

- Selección recurrente por habilidad combinatoria general.

- Selección recurrente por habilidad combinatoria específica.
- ii. Selección recurrente de familias de hermanos completos.
- iii. Selección recurrente de familias endocriadas:
 - Selección recurrente de familias S_1 (o $S_{0:1}$).
 - Selección recurrente de familias S_2 (o $S_{1:2}$).
 - Selección recurrente de familias homocigotas.
- iv. Selección recurrente combinada de familias de medios hermanos y S_1 .

Los métodos de selección recurrente interpoblacionales evalúan familias originadas por el cruzamiento entre componentes (individuos o familias) de dos poblaciones y se clasifican en:

- i. Selección recurrente interpoblacional de familias de medios hermanos.
Utilizando la población opuesta como probadora:
 - Selección recurrente interpoblacional de medios hermanos.
 - Selección recurrente interpoblacional modificada por Paterniani I y II.
 Utilizando una línea derivada de la población opuesta como probadora:
 - Selección recurrente interpoblacional de medios hermanos modificada.
- ii. Selección recurrente interpoblacional de familias de hermanos completos:
 - Selección recurrente interpoblacional de hermanos completos.
- iii. Selección recurrente recíproca modificada (Souza Jr., 1987; Figueiredo Santos *et al.*, 2007).

Los métodos de selección recurrente pueden combinarse y también alternarse entre ciclos de selección y cada uno de ellos, a su vez, presenta variantes según el tipo de familias que se utilizan para hacer la recombinación, el nivel de endocría de los individuos que generan las familias y el control parental. Estas particularidades, así como la magnitud de las evaluaciones en términos de número de ambientes de evaluación, número de repeticiones por ambiente, empleo de modelos mixtos, etc., constituyen un abanico de posibilidades de implementación y análisis que pueden afectar la respuesta a la selección, tanto medida por ciclo de selección como por unidad de tiempo.

10.3. Predicción de respuesta a diferentes métodos de selección recurrente

Considerando la ecuación de respuesta a la selección presentada en el Capítulo 9, la respuesta a la selección recurrente será, a igualdad de otros factores, proporcional a la heredabilidad. Todo lo que afecta el numerador de esta, es decir, la fracción o la totalidad de la variancia genética o aditiva de la población que está disponible para seleccionar, o al denominador, es decir, la variabilidad fenotípica modificará la respuesta. En cualquier caso, las decisiones del fitomejorador o los sucesos experimentales que alteren la covariancia entre las unidades de evaluación y los individuos que compondrán la población mejorada (C_M) modificará la respuesta esperada y la respuesta realizada.

Hallauer *et al.* (2010), citado por Márquez-Sánchez (2013), indicaron los niveles de respuesta promedios por ciclo de selección recurrente obtenidos en diferentes investigaciones. En selección recurrente espiga x hilera modificada fue 5,1 %, en selección de hermanos completos 5,8 %, en selección de líneas S_1 de 4,5 %. Más recientemente, la respuesta directa observada a la selección recurrente recíproca en doce estudios promedió 5,5 % por ciclo (Hallauer *et al.*, 2010).

Independientemente de ello, cuando el fitomejorador quiere comparar métodos alternativos de selección es importante que coteje las respuestas esperadas por unidad de tiempo, ya que los diferentes métodos insumen un número variable de estaciones de crecimiento, que pueden afectar o no el plazo temporal que demanda completar cada ciclo de selección recurrente. Ello dependerá del empleo de campos de invierno, disponibilidad de ambientes controlados para el avance de generaciones, etc.

Tanto Eberhart (1970) como Vencovsky (1980) y Vencovsky y Barriga (1992) han publicado sendos procedimientos que permiten predecir la respuesta obtenible por ciclo bajo distintos esquemas selectivos. Aplicando los conceptos presentados en el Capítulo 9, podemos determinar la ecuación de respuesta a los diferentes métodos de selección:

$$R = icrV_A / (VP)^{0,5},$$

donde:

i es la intensidad de selección,

c es el control parental,

$r = 2f_{xy}$ es el coeficiente de parentesco e indica la proporción de la variancia genética o aditiva que se expresa entre el tipo de familias empleadas como unidades de evaluación (ver Capítulo 6).

V_A es la estimación de la variancia aditiva de la población que se está seleccionando.

V_p es la estimación de la variancia fenotípica entre el tipo de familias o unidades de evaluación para el carácter que se está seleccionando.

El control parental admite dos valores posibles: $c=1,0$ cuando los individuos de la población mejorada solo resultan del cruzamiento de familias o individuos que han sido seleccionados. Alternativamente, $c=0,5$ si los individuos de la población mejorada resultan del cruzamiento entre individuos seleccionados (o sus familias) como progenitores hembra, polinizados por una muestra aleatoria de toda la población (que comprende tanto individuos seleccionados como no seleccionados).

La ecuación que sirve para estimar la respuesta a la selección presentada más arriba es la de uso corriente y asume que se está trabajando con poblaciones de tamaño infinito. Sin embargo, siendo rigurosos, cuando se selecciona una población de tamaño efectivo poblacional pequeño, la ecuación de predicción de respuesta debe modificarse (Souza Jr. *et al.*, 2000) para evitar sobreestimaciones. La expresión más aproximada es:

$$R = [icr(V_A + \Delta F.D_1)/(V_p)^{0.5}] - \Delta F.ID,$$
$$R = [ic'(\sigma_A^2 + \Delta F.D_1)/\sigma_p] - (\Delta F)(ID),$$

donde,

ΔF es el cambio en el coeficiente de endogamia de un ciclo al siguiente,

ID es la depresión por endocria y

D_1 es la covariancia entre los efectos aditivos y de dominancia de los homocigotas.

10.3.1. Procedimiento de predicción de respuesta de Eberhart

El procedimiento propuesto por Eberhart (1970) se basa en la siguiente expresión

$$R = ic'rV_A/(V_p)^{0.5},$$

donde:

i es la intensidad de selección,

c' es un coeficiente cuyo valor depende del control parental y de cuál es la unidad de recombinación,

$r = 2f_{xy}$ es el coeficiente que indica la fracción de la variancia genética o aditiva que se expresa entre el tipo de familias utilizadas como unidades de evaluación.

V_A es la variancia aditiva de la población que se está seleccionando.

V_p es la variancia fenotípica entre las unidades de evaluación para el carácter que se está seleccionando.

El coeficiente c' de Eberhart (1970) admite tres valores posibles:

- $c'=0,5$ cuando se utiliza como unidad de evaluación (X) y como unidad de recombinación (R) a un mismo tipo de familia, y la recombinación de individuos o familias seleccionados ocurre en un solo sexo. Dicho de otro modo, los individuos de la población mejorada surgen del cruzamiento entre familias seleccionadas y una muestra aleatoria de la totalidad de las familias evaluadas.
- $c'=1$ cuando las unidades de evaluación (X) y de recombinación (R) son el mismo tipo de familias y los individuos de la población mejorada (Y) son producto exclusivamente del cruzamiento de familias de individuos que han sido seleccionados.
- $c'=2$ cuando la unidad de evaluación (X) y de recombinación (R) son tipos de familias diferentes y los individuos de la población mejorada son producto exclusivamente del cruzamiento de familias de individuos que han sido seleccionados.

En el caso de la selección recurrente interpoblacional, la expresión de la respuesta a la selección es la suma de dos términos, cada uno correspondiente a la contribución de una y otra de las dos poblaciones de cría involucradas y en cada término el coeficiente es $c'=1$.

10.3.2. Procedimiento de predicción de respuesta de Vencovsky

El procedimiento empleado por Vencovsky (1980) y Vencovsky y Barriga (1992) utiliza para estimar la respuesta esperada a la siguiente expresión:

$$R = \theta V_A / (V_p)^{0,5},$$

donde:

V_A es la variancia aditiva de la población que se está seleccionando.

V_p = es la variancia fenotípica del tipo de familias que se evalúan.

θ es el producto de los coeficientes k entre cada paso de la genealogía que conecta la unidad de evaluación (X) con un individuo (Y) en la población mejorada, pasando por el individuo (P) de la población bajo selección, sumado a través de todos los senderos posibles que conectan X con Y. Los coeficientes de paso k pueden tomar uno de dos valores posibles:

- $k=0,5$ si el paso involucra un cruzamiento.
- $k=1,0$ si el paso involucra una autofecundación.

Incorporando una modificación al método de Vencovsky, este puede extenderse a otras situaciones en las que se evalúan familias resultantes de más de una autofecundación. En todos los casos, el coeficiente k es igual a $1 + F_t$, siendo F_t el coeficiente de endocría de la última filial que se autofecundó. Así cuando se hace selección entre familias $S_{0:1}$, $k=1+0$; cuando la selección es entre familias $S_{1:2}$, $k=1+0,5$. Para la selección entre familias $S_{2:3}$, $k=1+0,75$.

Los pasos de la genealogía que conecta la unidad de evaluación (X) con un individuo de la población mejorada (Y) se indican en el esquema con una línea simple, o con una línea doble, cuando involucran un cruzamiento o una autofecundación, respectivamente.

Tanto $c'r V_A$ (Eberhart) como θV_A (Vencovsky) cuantifican a la variancia genética aprovechada por el método de selección. La constante θ depende del control parental, de la fracción de la variancia genética expuesta por el tipo de familias evaluadas y de la modalidad de la recombinación.

Ejemplificaremos la aplicación de ambas metodologías en algunos métodos de selección recurrente. Gran parte de ellos fueron desarrollados para el mejoramiento de maíz, aunque son factibles de ser aplicados en otras especies cultivadas, de la manera original o con variantes.

10.4. Respuesta a la selección recurrente intrapoblacional

10.4.1. Selección recurrente fenotípica

Es el método de selección más simple, pero suficientemente efectivo para caracteres de alta heredabilidad. Los valores de mejora de los individuos de la población se estiman a partir de sus respectivos valores fenotípicos, y los mismos individuos seleccionados participan de la recombinación.

Siguiendo la nomenclatura anterior, el individuo de la población bajo selección es a su vez unidad de evaluación y unidad de recombinación. Cuando la selección se practica para un carácter cuya expresión fenotípica ocurre después de la floración, solo puede practicarse selección en uno de los sexos, de manera que el control parental es del 50 %. Es decir, que cualquier individuo de la población mejorada (Y_{CM}) habrá recibido el 50 % de sus genes de la planta madre que le dio origen y el resto proviene al azar, tanto de individuos seleccionados como no seleccionados. Si el carácter puede evaluarse previo a la floración, o polinización, es posible lograr el 100 % del control parental, tal como indica la Figura 10.3.

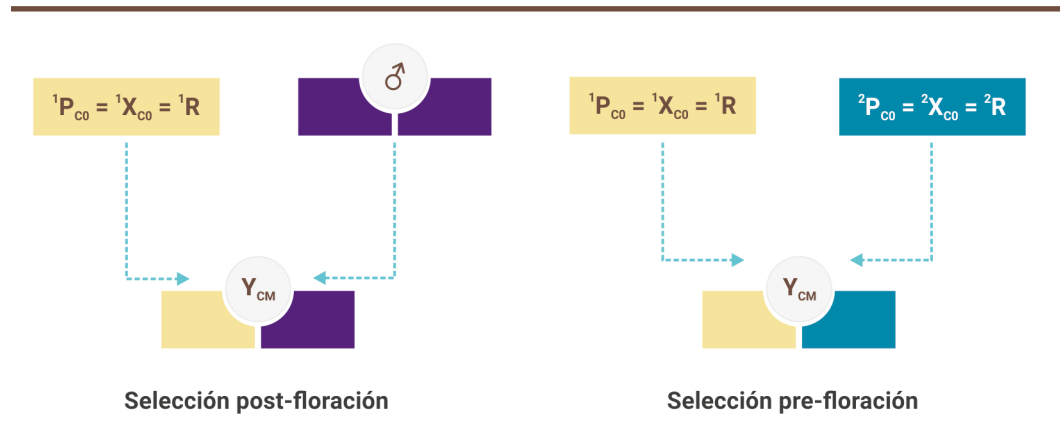


Figura 10.3. Esquema de la selección recurrente fenotípica con recombinación (ver texto).

Cuando la selección se practica en un solo sexo, la covarianza genética entre un individuo P seleccionado y un individuo Y en la población mejorada es equivalente a la covarianza Padre-Progenie, es decir, $(1/2)$ de la variancia aditiva en la población original. Si la selección es previa a la floración, el control parental es en ambos sexos, y por lo tanto el proceso selectivo puede aprovechar la totalidad de la variancia aditiva poblacional. Por lo tanto las estimaciones de respuesta a la selección serán:

$$R = i(1/2)(V_A)/(V_P)^{0.5} \text{ (selección entre individuos posterior a la floración).}$$

$$R = i(V_A)/(V_P)^{0.5} \text{ (selección entre individuos anterior a la floración).}$$

En el caso del procedimiento de Eberhart, el control parental define el tipo de parentesco entre P e Y, Progenitor-Progenie y Progenitor Medio-Progenie, por lo cual el término $c'r$ es $1/2$ y 1, respectivamente.

Al aplicar el método de Vencovsky:

$$1P_{co} = 1X_{co} = 1R \xrightarrow{1/2} Y_{CM}$$



$$R = i\theta V_A / (V_P)^{0.5} = i (1/2) V_A / (V_P)^{0.5}$$

$${}^1P_{CO} = {}^1X_{CO} = {}^1R \xrightarrow{1/2} Y_{CM} \xleftarrow{1/2} {}^2R = {}^2X_{CO} = {}^2P_{CO}$$

$$R = i\theta V_A / (V_P)^{0.5} = i [(1/2) + (1/2)] V_A / (V_P)^{0.5} = i V_A / (V_P)^{0.5}$$

Una de las debilidades de la selección recurrente fenotípica es el escaso control de la variabilidad ambiental en el sitio experimental donde se efectúa la selección. Para mejorar el control ambiental Gardner (1961) propuso dividir el sitio en parcelas o estratos más o menos pequeños. Al momento de seleccionar se procede a seleccionar igual porcentaje de individuos en cada estrato.

En el caso de la selección recurrente fenotípica sin estratificación, la variancia fenotípica es:

$$V_P = V_A + V_D + V_E$$

Donde V_E es la variabilidad ambiental (ver Capítulo 5)

Si hubiesen existido parcelas (estratos),

$$V_E = V_e + V_{ew}$$

donde V_e es la variabilidad ambiental entre parcelas y V_{ew} es la variabilidad ambiental dentro de parcelas o unidades experimentales.

En el caso de la selección recurrente fenotípica estratificada, la variancia fenotípica es:

$$V_P = V_A + V_D + V_E$$

Donde V_E es solo el componente de variabilidad ambiental dentro del estrato:

$$V_E = V_{ew}$$

Por lo tanto la variancia fenotípica, que incide en el denominador de la predicción de respuesta, si se estratifica resulta menor que sin estratificación.

10.4.2. Selección recurrente espiga x hilera (procedimiento original)

En este método (Hopkins, 1899) se cosechan separadamente un número (n) de espigas, cuyas semillas constituyen, las de cada espiga, n familias de medios hermanos. Al cosechar estas espigas en los n individuos de la población original (${}^1P_{CO}, {}^2P_{CO} \dots {}^n P_{CO}$), solo conocemos al progenitor femenino (precisamente ${}^1P_{CO}$) de esas familias de medios hermanos. La gameta mas-

culina provino al azar de la población. Utilizaremos la semilla cosechada de cada espiga para hacer la evaluación de esas familias de medios hermanos (X) en un experimento con características particulares: se sembrará una hilera, surco, o parcela con la semilla de cada una de las n familias en condiciones de aislamiento de fuentes de polen extraño proveniente de otras poblaciones de maíz. Al mismo tiempo se hará una mezcla mecánica de semilla (pozo o *bulk* B) con igual cantidad de semilla de cada una de las n familias para evaluar. Con ella se sembrarán surcos intercalados entre las n familias para evaluar. Las unidades de evaluación (X) solo tienen un 50 % de la información genética del individuo $^1P_{CO}$. El *bulk* B se utilizará como polinizador común de todas las familias de medios hermanos, para lo cual los surcos de medios hermanos deberán ser despanojados antes de su floración. A la cosecha se registrarán los pesos de semilla (si ese fuera el carácter de interés) y con esa información se seleccionará un pequeño grupo, normalmente 10-20 % de la muestra de individuos de la población. Obsérvese que las semillas en las espigas que se cosechen sobre los surcos seleccionados producirán los individuos (Y) de la población mejorada y son producto del cruzamiento de las familias por una muestra aleatoria de polen (de plantas pertenecientes a familias que fueron seleccionadas y otras que no) de la población original C_0 .

Como vemos, las unidades de evaluación (X) y recombinación (R) pertenecen a un mismo tipo de familias (medios hermanos) y el control parental es sobre un solo sexo ($c=0,5$). El esquema correspondiente se muestra en la Figura 10.4.

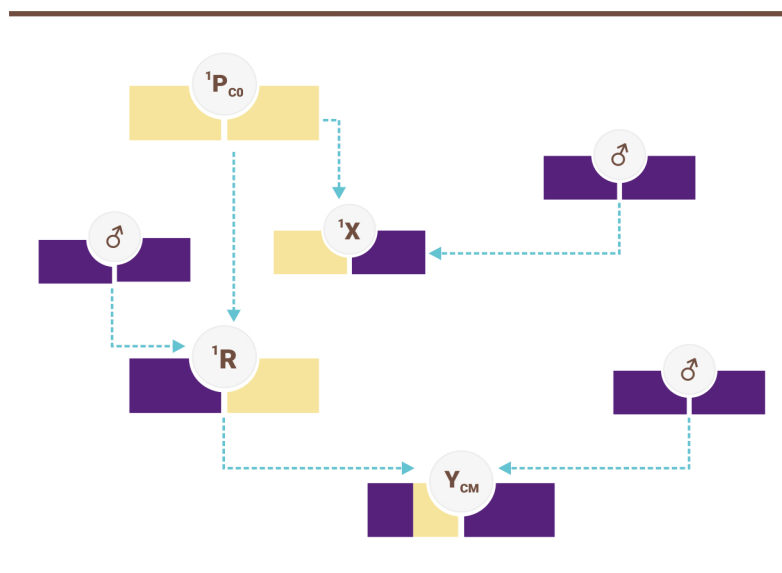


Figura 10.4.
Esquema de la selección recurrente espiga x hilera (original) (ver texto).

Para este esquema los procedimientos de Eberhart y de Vencovsky dan, respectivamente, las siguientes expresiones:

$$R = i(c')(rV_A)/(V_P)^{0.5} = i(1/2)(1/4V_A)/(V_P)^{0.5} = i(1/8)V_A/(V_P)^{0.5}$$

$$X^1 \xrightarrow{1/2} P_{C0}^1 \xrightarrow{1/2} R_1 \xrightarrow{1/2} Y_{CM}^1$$

$$R = i\theta V_A/(V_P)^{0.5} = i(1/2)(1/2)(1/2)V_A/(V_P)^{0.5} = i(1/8)V_A/(V_P)^{0.5}$$

10.4.3. Selección recurrente espiga x hilera modificada

La selección espiga por hilera modificada (Lonnquist, 1964) mejora en gran medida la respuesta esperable a la selección respecto del método original porque se ejerce un mayor control de la variancia fenotípica al reducir las contribuciones de la variancia del error y de la variancia aditiva x ambiente a la variancia fenotípica, debido a que los experimentos se hacen con repeticiones y en más de un ambiente de evaluación. En este método las familias de medios hermanos se obtienen de la misma manera y el control parental es similar al del procedimiento original. Las unidades de evaluación y de recombinación son coincidentes. La ventaja de este método estriba en poder seleccionar con una mayor heredabilidad respecto al procedimiento de Hopkins por reducción de la variancia fenotípica.

El diagrama de la selección es semejante al de la selección espiga x hilera original.

La aplicación de los procedimientos de Eberhart y de Vencovsky para predecir la respuesta a este esquema particular de selección entre familias de medios hermanos da respectivamente las siguientes expresiones:

$$R = i(c')(rV_A)/(V_P)^{0.5} = i(1/2)(1/4V_A)/(V_P)^{0.5} = i(1/8)V_A/(1/4V_A + 1/4V_{AxE}/a + V_{EE}/ar)^{0.5}$$

$$R = i\theta V_A/(V_P)^{0.5} = i[(1/2)(1/2)(1/2)]V_A/(V_P)^{0.5} = i(1/8)V_A/(1/4V_A + 1/4V_{AxE}/a + V_{EE}/ar)^{0.5}$$

Cabe mencionar aquí que el método de selección espiga por hilera original y el de espiga x hilera modificada también contempla la selección dentro de familias de medios hermanos, por lo cual la respuesta total por ciclo de selección debe considerar también la respuesta esperada a la selección dentro de las familias. Por lo tanto, la respuesta total (R_T) esperada a la selección de espiga x hilera y espiga por hilera modificada será:

$$R_T = R + R_{DF}$$

donde R_{DF} es la respuesta esperada a la selección dentro de familias de medios hermanos:

$$R_{DF} = ic'(rV_A)/(V_P)^{0.5} = i(1/2)(3/4)V_A / [(3/4)V_A + V_D + (3/4)V_{AE} + V_{DE} + V_{ew}]^{0.5}$$

10.4.4. Selección recurrente espiga x hilera doble modificada

La selección espiga por hilera doble modificada (Compton y Comstock, 1976) difiere del esquema propuesto por Lonquist y permite aumentar la respuesta por ciclo de selección al mejorar el control parental ya que los individuos de la población mejorada (Y) se generarán únicamente por cruzamiento entre familias seleccionadas. El cumplimiento de un ciclo de selección con este método demanda una estación de crecimiento adicional puesto que se utilizará para la recombinación solo a la semilla remanente de las familias de medios hermanos que resultaron seleccionadas. El diagrama de esta selección se presenta en la Figura 10.5.

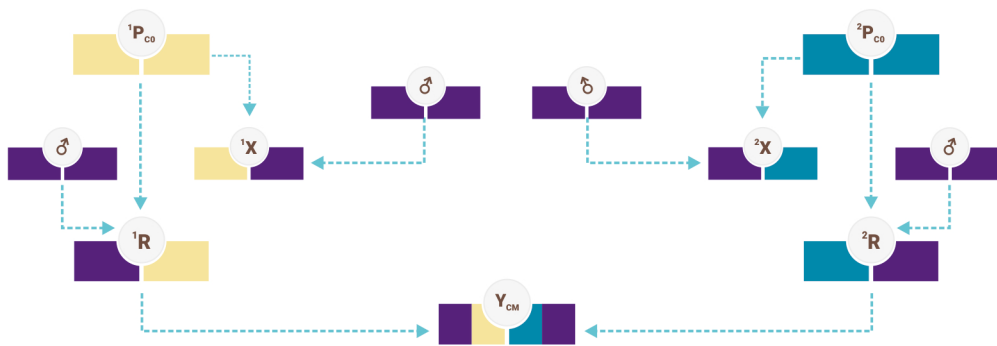


Figura 10.5. Esquema de la selección recurrente espiga x hilera doble modificada (ver texto).

Para este esquema de selección los procedimientos de Eberhart y de Venkovsky dan respectivamente las siguientes expresiones:

$$R = i(c')(rV_A)/(V_P)^{0.5} = i(1)(1/4V_A)/(V_P)^{0.5} = i(1/4)V_A / (1/4V_A + 1/4V_{AxE}/a + V_{EE}/ar)^{0.5}$$

$$X^1 \xrightarrow{1/2} P^1_{CO} \xrightarrow{1/2} R_1 \xrightarrow{1/2} Y_{CM} \xleftarrow{1/2} R^2 \xleftarrow{1/2} P^2_{CO} \xleftarrow{1/2} X^2$$

$$R = i\theta V_A / (V_P)^{0.5} = i [(1/2)(1/2)(1/2) + (1/2)(1/2)(1/2)] V_A / (V_P)^{0.5}$$

$$= i(1/8 + 1/8)V_A / (1/4V_A + 1/4V_{AxE}/a + V_{EE}/ar)^{0.5} = i(1/4)V_A / (1/4V_A + 1/4V_{AxE}/a + V_{EE}/ar)^{0.5}$$

10.4.5. Otros métodos de selección de medios hermanos

Selección recurrente entre cruzamientos de prueba (Selección recurrente por habilidad combinatoria general y selección recurrente por habilidad combinatoria específica)

La selección recurrente por habilidad combinatoria general (Jenkins, 1940) y la selección por habilidad combinatoria específica (Hull, 1945) emplean también familias de medios hermanos porque evalúan cruzamientos de prueba de los individuos (P) de la población original contra un genotipo probador común. La denominación de estos métodos de selección por tipo de habilidad combinatoria ha caído en desuso con el tiempo. Ambos métodos fueron concebidos para generar información sobre el tipo de acción génica involucrada en el control del rendimiento de grano en maíz y la denominación se relacionaba con las características del probador que utiliza cada uno. En la selección recurrente por habilidad combinatoria general, el probador es genéticamente heterogéneo y de base genética relativamente amplia. Puede tratarse de la propia población, otra diferente, o de un híbrido doble. En la selección recurrente por habilidad combinatoria específica, el probador es genéticamente homogéneo y de base genética más estrecha, por ejemplo un híbrido simple o una línea endocriada. Por regla general, en los programas de mejoramiento de especies alógamas donde la dominancia es importante, prefiere utilizarse un probador que pertenezca a un grupo heterótico diferente, a fin de acoplar la actividad de mejoramiento de la población con la de desarrollo (y evaluación) de líneas e híbridos superiores, economizando tiempo y otros recursos.

En ambos métodos, con los individuos de la población original (P) se generan dos tipos de familias: las de medios hermanos (cruzamientos de prueba) que se utilizarán como unidades de evaluación (X) y las familias de autofecundación ($S_{0:1}$) que se utilizarán como unidades de recombinación (R). Los individuos de la población mejorada (Y) se generan por cruzamiento entre familias ($S_{0:1}$) correspondientes a los individuos (P) seleccionados.

La decisión sobre el probador para utilizar es particularmente crítica. Primeramente por lo señalado en capítulos previos, pero también por consideraciones prácticas. Cuando los cruzamientos de prueba se hacen utilizando un probador heterogéneo es importante muestrear adecuadamente al probador por cuanto hay que asegurarse de que realmente el probador haya sido “común” a todos los individuos de la población original que se están evaluando. Cuando el probador es homogéneo ese problema de muestreo no existe y el número de individuos que se cruzan solo debe ser el suficiente para disponer de la semilla del cruzamiento en cantidad necesaria para realizar los experimentos de evaluación. Toda vez que se utilice un probador diferente a la propia población, la variancia entre cruzamientos

de prueba no resulta ser una función lineal de las variancias genéticas de la población original, a menos que se cumplan determinados supuestos.

El esquema o diagrama general de ambos métodos de selección recurrente de medios hermanos se presenta en la Figura 10.6.

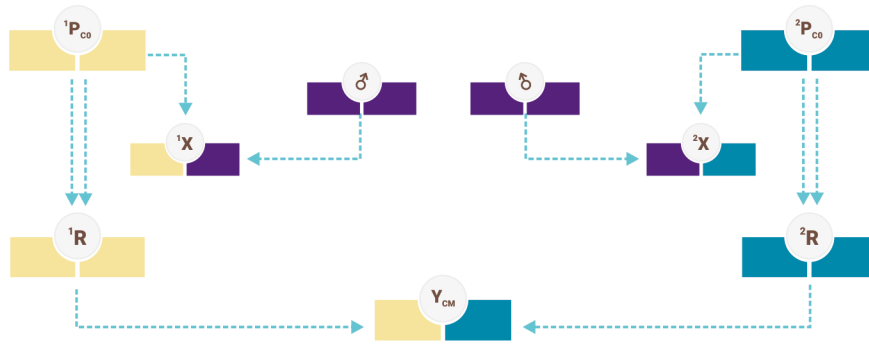
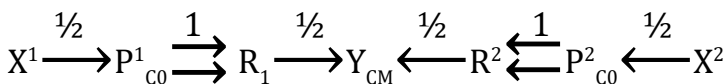


Figura 10.6. Esquema de la selección recurrente de medios hermanos por habilidad combinatoria (ver texto).

La aplicación de los procedimientos de Eberhart y de Vencovsky de predicción de respuesta a estos esquemas de selección entre familias de medios hermanos dan respectivamente las siguientes expresiones:

$$R = i(c')(rV_A)/(V_P)^{0.5} = i(2)(1/4)V_A/(V_P)^{0.5} = i(1/2)V_A/((1/4V_A + 1/4V_{AxE}/a + V_{EE}/ar)^{0.5}$$



$$R = i\theta V_A/(V_P)^{0.5} = i[(1/2)(1)(1/2) + (1/2)(1)(1/2)]V_A/(V_P)^{0.5}$$

$$= i(1/4 + 1/4)V_A/(1/4V_A + 1/4V_{AxE}/a + V_{EE}/ar)^{0.5}$$

$$= i(1/2)V_A/(1/4V_A + 1/4V_{AxE}/a + V_{EE}/ar)^{0.5}$$

Hasta aquí lo presentado se aplica cuando el probador es la propia población. Una expresión generalizada de la variancia genética entre cruzamientos de prueba (*test crosses*) es:

$${}^{TC}\sigma_A^2 = pq[a + (s - r)d]^2/2, \quad [10.1]$$

donde

s y r son las frecuencias génicas del probador correspondientes a los loci cuyas frecuencias son q y p en la población original.

Si el probador fuera la propia población, la expresión anterior resulta en la expresión ya conocida:

$${}^{\text{TC}}\sigma_A^2 = pq[a + (q - p)d]^2/2.$$

Una expresión similar a [10.1] es la provista por Rawling y Thompson (1962) para niveles arbitrarios de endocria en el material en evaluación. Reemplazando s por $(1-r)$:

$${}^{\text{TC}}\sigma_A^2 = (1/2)pq(1 + F)[a + (1 - 2r)d]^2,$$

donde r es la frecuencia del alelo favorable en la población probadora.

Cuando existe dominancia, el factor crítico en la elección del probador puede deducirse a partir de esta última expresión. Si $d=0$, el efecto de las frecuencias génicas del probador es nula y la decisión sobre cuál es el probador más adecuado resulta un asunto trivial. Si $d \neq 0$, claramente el empleo de un probador cuya frecuencia del alelo favorable sea muy baja o próxima a cero producirá un incremento importante en la variancia de los cruzamientos de prueba, mejorando el poder discriminatorio de la evaluación (en última instancia la heredabilidad) y con ello la respuesta esperable de la selección (Hallauer, 1975). En los programas de mejoramiento destinados al desarrollo de híbridos este efecto del probador se logra escogiendo uno que pertenezca a un grupo heterótico opuesto al de la población que se pretende mejorar, aumentando así la probabilidad de que la diferencia de frecuencias alélicas entre probador y material probado sea mayor y lo mismo ocurra con la variancia genética entre los cruzamientos de prueba.

Otro aspecto muy importante que se desprende de la expresión de Rawling y Thompson (1962) es que sería posible incrementar el numerador de la respuesta a la selección efectuando los cruzamientos con el probador, no ya de plantas S_0 de C_0 , sino de plantas con $F > 0$, por cuanto ello aumentaría la variancia entre cruzamientos de prueba. Por ejemplo, si en lugar de cruzar con el probador a plantas individuales S_0 de C_0 , se cruzaran plantas individuales que son, por ejemplo, $S_{0:1}$, o $S_{1:2}$, o $S_{0:2}$. En tales casos, las variancias entre cruzamientos de prueba serán, respectivamente:

$${}^{\text{TCS0:1}}\sigma_A^2 = (1/2)pq(1 + F)[a + (1 - 2r)d]^2 = (1/2)pq(1 + 1/2)[a + (1 - 2r)d]^2$$

$${}^{\text{TCS1:2}}\sigma_A^2 = (1/2)pq(1 + F)[a + (1 - 2r)d]^2 = (1/2)pq(1 + 3/4)[a + (1 - 2r)d]^2$$

$${}^{\text{TCS0:2}}\sigma_A^2 = (1/2)pq(1 + F)[a + (1 - 2r)d]^2 = (1/2)pq(1 + 1/2)[a + (1 - 2r)d]^2$$

Siguiendo a Empig *et al.* (1972), la fórmula de predicción de respuesta a la selección de cruzamientos de prueba con otra población utilizada como probadora o tester es:

$$R = i (1/2) pq ({}^T\alpha_i) ({}^{C0}\alpha_i) / (V_p)^{0.5},$$

Donde ${}^T\alpha_i$ y ${}^{C0}\alpha_i$ son los efectos promedio de sustitución de un gen en la población probadora y en la población original, respectivamente.

10.4.6. Selección recurrente de hermanos completos

En la selección recurrente de hermanos completos las progenies se forman por cruzamientos entre plantas o individuos tomados de a pares. Toda la semilla obtenida del cruzamiento entre cada par de individuos constituye una familia de hermanos completos, la que se utiliza con dos finalidades: para evaluación (X) y para la recombinación (R) en el caso de que ese par de plantas resultara seleccionado. Por lo tanto la variancia genotípica total que se expresa entre unidades de evaluación será $1/2V_A + 1/4V_D$. El control parental es $c'=1$ porque las unidades de evaluación y recombinación son las mismas y los individuos de la población mejorada proceden únicamente de la fusión de gametas de individuos seleccionados en la población original.

El esquema de selección entre hermanos completos se presenta en la Figura 10.7.

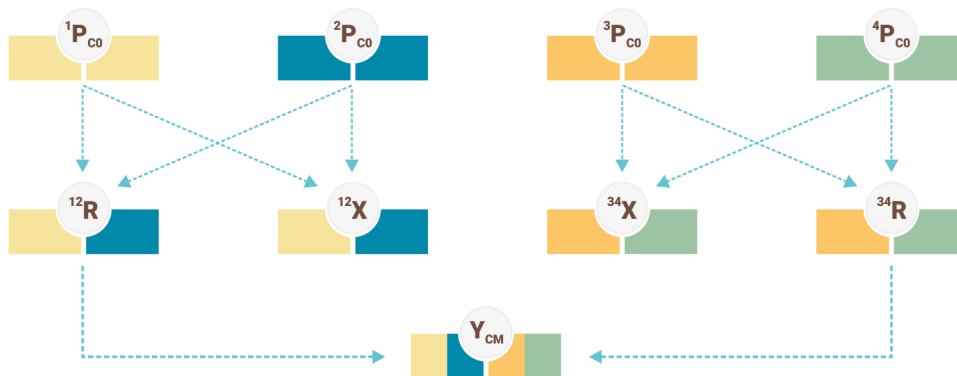


Figura 10.7. Esquema de la selección recurrente de hermanos completos (ver texto).

En este esquema de selección la aplicación de los procedimientos de Eberhart y de Vencovsky para predecir la respuesta dan respectivamente las siguientes expresiones:

$$R = i(c')(rV_A)/(V_p)^{0.5} = i(1) (1/2 V_A)/(1/2 V_A + 1/4 V_D + (1/2 V_{AxE} + 1/4 V_{DxE})/a + V_{EE}/ar]^{0.5}$$

$$X^{12} \xrightarrow{1/2} P_{C0}^1 \xrightarrow{1/2} R^{12} \xrightarrow{1/2} Y_{CM} \xleftarrow{1/2} R^{12} \xleftarrow{1/2} P_{C0}^2 \xleftarrow{1/2} X^{12}$$

$$X^{12} \xrightarrow{1/2} P_{C0}^2 \xrightarrow{1/2} R^{12} \xrightarrow{1/2} Y_{CM} \xleftarrow{1/2} R^{12} \xleftarrow{1/2} P_{C0}^2 \xleftarrow{1/2} X^{12}$$

$$X^{34} \xrightarrow{1/2} P_{C0}^3 \xrightarrow{1/2} R^{34} \xrightarrow{1/2} Y_{CM} \xleftarrow{1/2} R^{34} \xleftarrow{1/2} P_{C0}^3 \xleftarrow{1/2} X^{34}$$

$$X^{34} \xrightarrow{1/2} P_{C0}^4 \xrightarrow{1/2} R^{34} \xrightarrow{1/2} Y_{CM} \xleftarrow{1/2} R^{34} \xleftarrow{1/2} P_{C0}^4 \xleftarrow{1/2} X^{34}$$

$$R = i \theta V_A / (V_p)^{0.5}$$

$$= i [(1/2)(1/2)(1/2) + (1/2)(1/2)(1/2) + (1/2)(1/2)(1/2) + (1/2)(1/2)(1/2)] V_A / (V_p)^{0.5} =$$

$$i [4(1/8) V_A / (V_p)^{0.5} = i(1/2) V_A / (1/2 V_A + 1/4 V_D + (1/2 V_{AxE} + 1/4 V_{DxE})/a + V_{EE}/ar]^{0.5}$$

10.4.7. Selección recurrente de familias endocriadas $S_{0:1}$

En este método se autofecundan plantas de la población original y las semillas obtenidas (familias $S_{0:1}$) se utilizan como unidades de evaluación (X) y de recombinación (R). Solo se utiliza para la recombinación la semilla de autofecundación ($S_{0:1}$) de los individuos que resultaron seleccionados.

El esquema del método se presenta en la Figura 10.8.

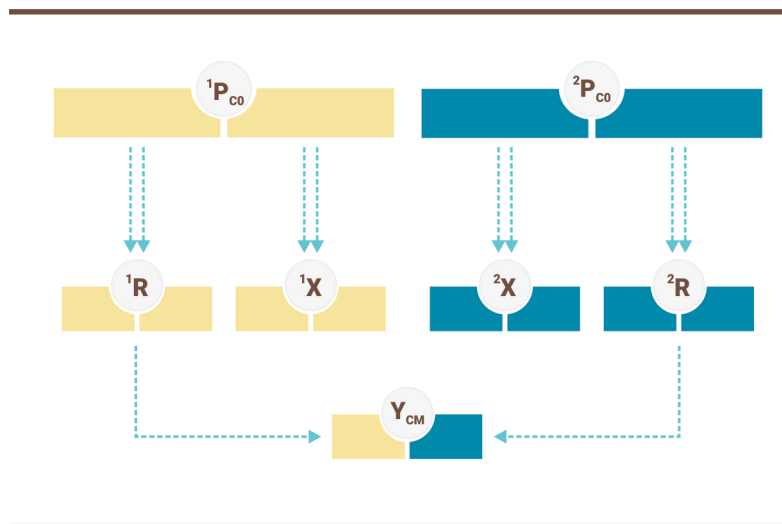
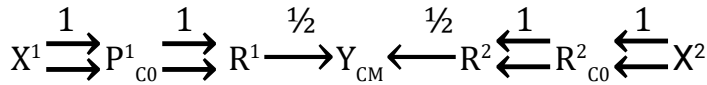


Figura 10.8. Esquema de la selección recurrente de progenies de autofecundación (ver texto).

En este esquema de selección, la aplicación de los procedimientos de Eberhart y de Vencovsky para predecir la respuesta dan respectivamente las siguientes expresiones:

$$R = i(c')(rV_A)/(V_P)^{0.5} = i(1)(1V_A)/(1V_A + 1/4V_D + (V_{AxE} + 1/4V_{DxE})/a + V_{EE}/ar]^{0.5}$$

$$= iV_A/(V_A + 1/4V_D + (V_{AxE} + 1/4V_{DxE})/a + V_{EE}/ar]^{0.5}$$



$$R = i\theta V_A/(V_P)^{0.5} =$$

$$= i[(1)(1)(1/2) + (1)(1)(1/2)]V_A/(1V_A + 1/4V_D + (1V_{AxE} + 1/4V_{DxE})/a + V_{EE}/ar]^{0.5}$$

$$= iV_A/(V_A + 1/4V_D + (V_{AxE} + 1/4V_{DxE})/a + V_{EE}/ar]^{0.5}$$

Debe tenerse en cuenta que la variancia entre familias $S_{0:1}$ no solo contiene la variancia aditiva, sino un término adicional diferente de cero, a menos que la población original posea frecuencias génicas intermedias ($p=q=0,5$). Por tal motivo la predicción de respuesta a la selección podría resultar sobrestimada (ver Capítulo 5).

10.4.8. Selección recurrente de familias endocriadas $S_{1:2}$

En este método se autofecundan plantas de la población original y sobre las filiales $S_{0:1}$ obtenidas se autofecundan plantas individuales y se genera semilla de familias $S_{1:2}$. Estas semillas se utilizan como unidad de evaluación y unidad de recombinación. En este último caso se empleará semilla de autofecundación ($S_{1:2}$) de los individuos que resultaron seleccionados.

El esquema del método se presenta en la Figura 10.9.

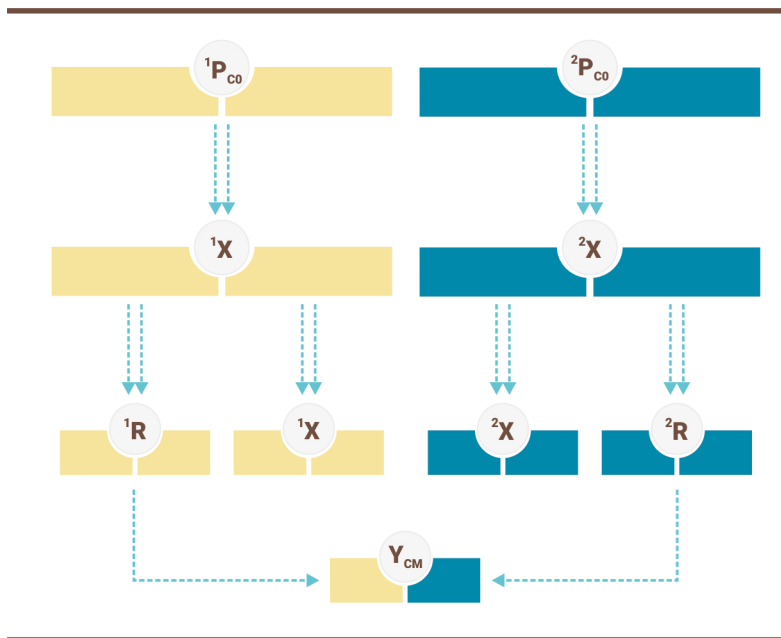
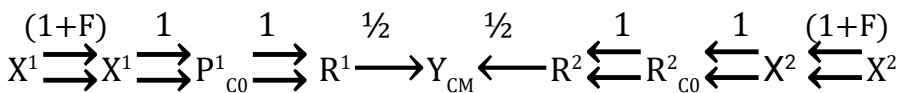


Figura 10.9. Esquema de la selección recurrente de progenies de autofecundación $S_{1,2}$ (ver texto).

En este esquema de selección, la aplicación de los procedimientos de Eberhart y de Vencovsky para predecir la respuesta dan, respectivamente, las siguientes expresiones:

$$R = i(c')(rV_A)/(V_p)^{0.5} =$$

$$= i(1)(3/2V_A)/\{(3/2)V_A + (3/16)V_D + [(3/2)V_{AxE} + (3/16)V_{DxE}]/a + V_{EE}/ar\}^{0.5}$$



$$R = i\theta V_A/(V_p)^{0.5} = i[(1)(1,5)(1/2) + (1)(1,5)(1/2)]V_A/(1,5V_A + (3/16)V_D +$$

$$+ [(1,5V_{AxE} + (3/16)V_{DxE}]/a + V_{EE}/ar]^{0.5}$$

$$= i(1,5)V_A/(1,5V_A + (3/16)V_D + (1,5V_{AxE} + 1/4 V_{DxE})/a + V_{EE}/ar]^{0.5}$$

10.5. Respuesta a la selección recurrente recíproca (interpoblacional)

Como hemos visto al inicio de este capítulo, el objetivo de la selección interpoblacional es mejorar el comportamiento del cruzamiento entre dos poblaciones. Miranda Filho (1982) y Miranda Filho y Paterniani (1983) establecieron que la selección interpoblacional se justifica respecto de la

intrapoblacional de las dos poblaciones separadamente, cuando el parámetro Φ resulta mayor que 1, definiéndolo como:

$$\Phi = (\sigma_{A12}^2 + \sigma_{A21}^2) / (\sigma_{A1A2} + \sigma_{A2A1})$$

Donde el numerador se refiere a la suma de las dos variancias genéticas de los cruzamientos interpoblaciones y el denominador es la suma de las covariancias entre los valores aditivos entre ambas poblaciones.

El cambio esperable en la heterosis por efecto de la selección interpoblacional (Souza Jr. y Miranda Filho, 1985) será:

$$\Delta h \approx 2(p - r)d(\Delta p - \Delta r)$$

Obsérvese que el aumento de heterosis será mayor en la medida en que las dos poblaciones parentales vayan diferenciándose en frecuencias génicas por efecto de la selección.

10.5.1. Selección recurrente de medios hermanos interpoblacionales o selección recurrente recíproca de medios hermanos

Este procedimiento desarrollado por Comstock *et al.* (1949) consiste en generar dos tipos de progenies de individuos tomados en ambas poblaciones, A y B. Los individuos S_0 de A se autofecundan para generar familias ($^A S_{0:1}$) y a su vez se cruzan con un grupo de plantas de la población B para generar familias de medios hermanos interpoblacionales A x B ($^A M H_{AxB}$). Lo mismo se realiza en la población B generando familias ($^B S_{0:1}$) y ($^B M H_{BxA}$). Las familias interpoblacionales serán las unidades de evaluación (X) que se probarán en sendos experimentos: el correspondiente a $^A M H_{AxB}$ y el correspondiente a $^B M H_{BxA}$. Las semillas de autofecundación, $^A S_{0:1}$ y $^B S_{0:1}$, se emplearán como unidades de recombinación para generar las poblaciones $^A C_M$ y $^B C_M$.

En la Figura 10.10 se presenta el esquema de la selección recurrente recíproca de medios hermanos.

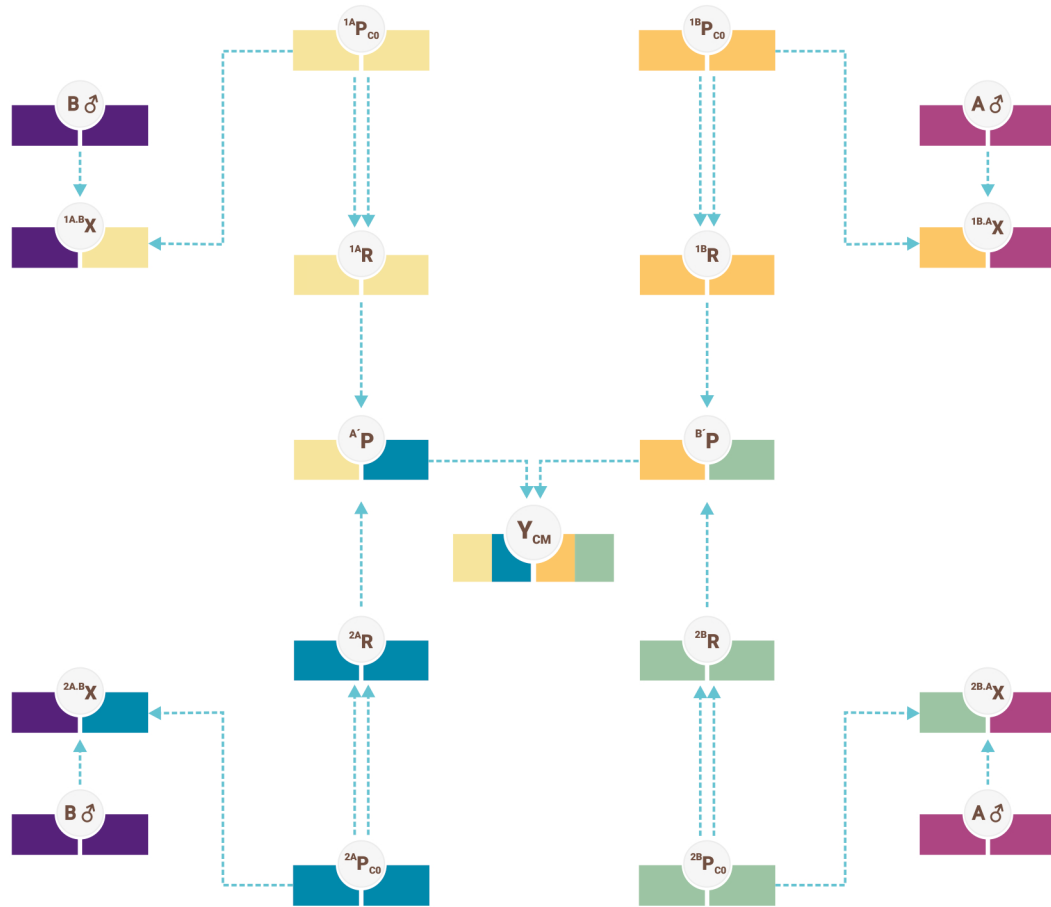


Figura 10.10. Esquema de la selección recurrente recíproca de medios hermanos (ver texto).

De modo análogo al caso de la variancia genotípica intrapoblacional, en la interpoblacional existirán los dos componentes causales de esta, correspondientes a la variancia aditiva y la variancia de dominancia interpoblacionales:

$${}^{AxB}\sigma_G^2 = {}^{AxB}\sigma_A^2 + {}^{AxB}\sigma_D^2 = (1/2)({}^{AxB}\sigma_A^2 + {}^{BxA}\sigma_A^2) + (1/2)({}^{AxB}\sigma_D^2 + {}^{BxA}\sigma_D^2)$$

Tratándose de familias de medios hermanos interpoblacionales, la variancia de dominancia que se expresa entre ellas es cero.

Si las frecuencias génicas en las poblaciones A y B fueran iguales,

$${}^A\sigma_{(AxB)}^2 = {}^A\sigma_{(BxA)}^2 = \sigma_A^2$$

En este esquema de selección, la aplicación de los procedimientos de Eberhart y de Vencovsky para predecir la respuesta dan respectivamente las siguientes expresiones:

$$R = i(c')(rV_A)/(V_P)^{0.5} = i\{ (1)(1/4)^{AxBV_A}/[(1/4)^{AxBV_A} + (1/4)^{AxBV_{AxE}}/a + V_{EE}/ar]^{0.5} + (1)(1/4)^{BxA V_A}/[(1/4)^{BxA V_A} + (1/4)^{BxA V_{AxE}}/a + V_{EE}/ar]^{0.5} \}$$

$$R = i\theta V_A/(V_P)^{0.5} = i\{ [(1/2)(1)(1/2)(1/2) + (1/2)(1)(1/2)(1/2)]^{AxBV_A}/[(1/4)^{AxBV_A} + (1/4)^{AxBV_{AxE}}/a + V_{EE}/ar]^{0.5} + [(1/2)(1)(1/2)(1/2) + (1/2)(1)(1/2)(1/2)]^{BxA V_A}/[(1/4)^{BxA V_A} + (1/4)^{BxA V_{AxE}}/a + V_{EE}/ar]^{0.5} \}$$

$$= i(1/4)^{AxBV_A}/[(1/4)^{AxBV_A} + (1/4)^{AxBV_{AxE}}/a + V_{EE}/ar]^{0.5} + i(1/4)^{BxA V_A}/[(1/4)^{BxA V_A} + (1/4)^{BxA V_{AxE}}/a + V_{EE}/ar]^{0.5}$$

Recordemos que en los casos de selección interpoblacional, cuando se aplica la regla de Eberhart, $c' = 2$ por cuanto las unidades de recombinación y selección o evaluación son diferentes, pero se “reparte” en ambos términos de la suma de contribuciones de cada población a la respuesta final.

Si ambas poblaciones tuvieran las mismas frecuencias génicas, la respuesta esperada a la selección resultaría igual a la de la selección recurrente intrapoblacional de medios hermanos con recombinación de familias $S_{0:1}$.

Paterniani y Vencovsky (1977) y Paterniani (1978) propusieron modificaciones al procedimiento original de Comstock *et al.* (1949) para reducir la magnitud de trabajo manual que implica hacer las familias interpopulacionales mediante polinizaciones controladas. En una de las propuestas se siembra una muestra de familias de medios hermanos de A en un lote aislado por distancia o fecha de otras poblaciones de maíz (para evitar contaminación con polen extraño) y en otro lote se siembra una muestra de familias de medios hermanos de la población B, también en condiciones de aislamiento. Para hacer los cruzamientos interpopulacionales, en el lote donde se sembraron las familias de medios hermanos de A se las despanoja antes de la floración y se emplean como hembras para ser polinizadas por un macho compuesto de un *bulk* de individuos de familias de medios hermanos de B. En el otro lote, las familias de medios hermanos de B se siembran para ser cruzadas por un *bulk* de familias de medios hermanos de A. Una fracción de la semilla de las familias de medios hermanos de A y de B, que se sembraron en sendos lotes para hacer los cruzamientos interpopulacionales, se deja en reserva para utilizarla como unidad de recombinación (R) en el caso de las familias seleccionadas. Las semillas de cruzamientos interpopulacionales se emplearán como unidades de evaluación (X).

El esquema de esta variante se presenta en la Figura 10.11.

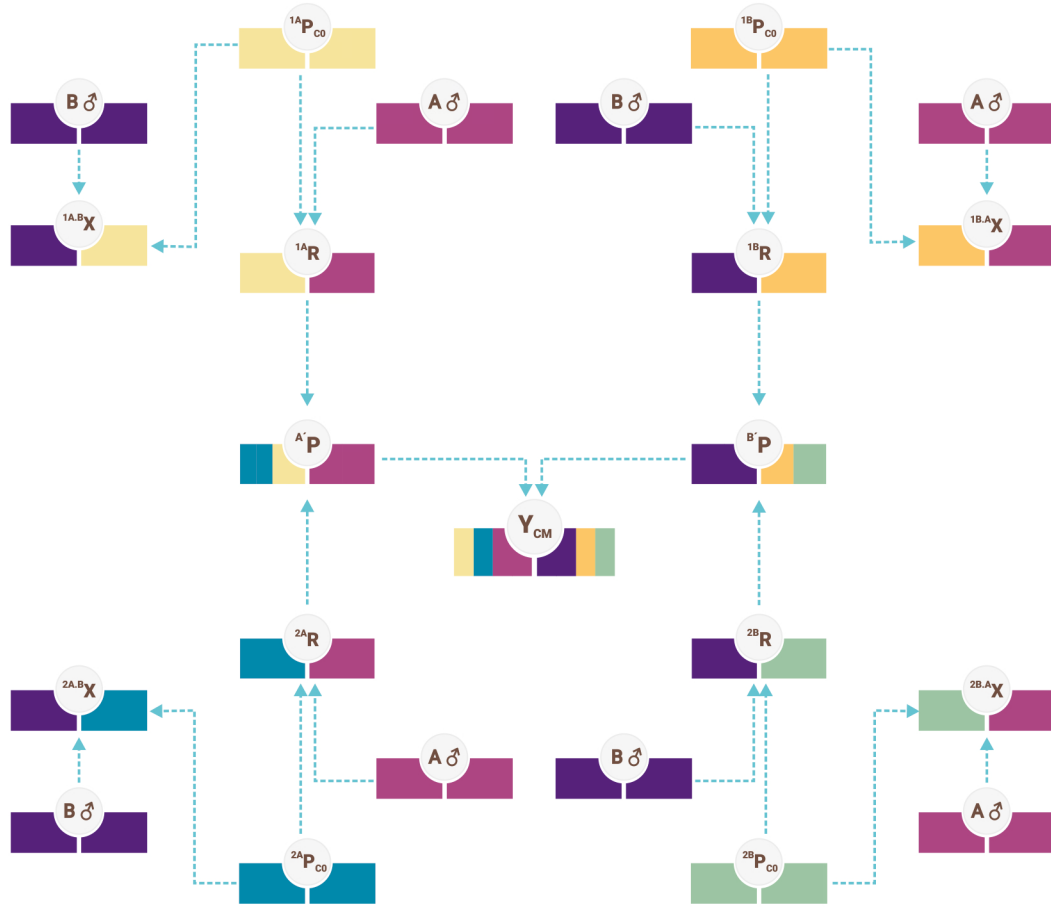


Figura 10.11. Esquema de la selección recíproca de medios hermanos. Variante de Paterniani (ver texto).

Para este esquema de selección, la aplicación de los procedimientos de Eberhart y de Vencovsky para predecir la respuesta dan respectivamente las siguientes expresiones:

$$R = i(c')(rV_A)/(V_p)^{0.5} = i\left\{\frac{1}{2}\right\}\left\{\frac{1}{4}\right\}^{AxBV_A}/\left[\left(\frac{1}{4}\right)^{AxBV_A} + \left(\frac{1}{4}\right)^{AxBV_{AxE}}/a + V_{EE}/ar\right]^{0.5} + \left\{\frac{1}{2}\right\}\left\{\frac{1}{4}\right\}^{BxAV_A}/\left[\left(\frac{1}{4}\right)^{BxAV_A} + \left(\frac{1}{4}\right)^{BxAV_{AxE}}/a + V_{EE}/ar\right]^{0.5}\right\}$$

$$R = i\theta V_A/(V_p)^{0.5} = i\left\{\left[\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\right]^{AxBV_A}/\left[\left(\frac{1}{4}\right)^{AxBV_A} + \left(\frac{1}{4}\right)^{AxBV_{AxE}}/a + V_{EE}/ar\right]^{0.5} + \left[\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\right]^{BxAV_A}/\left[\left(\frac{1}{4}\right)^{BxAV_A} + \left(\frac{1}{4}\right)^{BxAV_{AxE}}/a + V_{EE}/ar\right]^{0.5}\right\}$$

$$= i\left\{\frac{1}{8}\right\}^{AxBV_A}/\left[\left(\frac{1}{8}\right)^{AxBV_A} + \left(\frac{1}{8}\right)^{AxBV_{AxE}}/a + V_{EE}/ar\right]^{0.5} + i\left\{\frac{1}{8}\right\}^{BxAV_A}/\left[\left(\frac{1}{8}\right)^{BxAV_A} + \left(\frac{1}{8}\right)^{BxAV_{AxE}}/a + V_{EE}/ar\right]^{0.5}\right\}$$

En este caso de selección interpoblacional, cuando se aplica la regla de Eberhart, $c=1$ por cuanto las unidades de recombinación y selección o evaluación son iguales, pero se “reparte” en ambos términos de la suma de contribuciones de cada población a la respuesta final. Por su parte, Marquez Sánchez (1987) propuso modificaciones al procedimiento alternativo de Paterniani y Vencovsky.

10.5.2. Selección recurrente recíproca de hermanos completos

El método originalmente propuesto requiere el cruzamiento de pares de plantas prolíficas S_0 de las poblaciones A y B (Hallauer y Eberhart, 1970; Hallauer, 1984, 1999). Sobre las mismas plantas en la espiga restante se hace autofecundación. Por lo tanto sobre cada par de plantas cruzadas se obtendrán dos tipos de progenies: la semilla de hermanos completos debida al cruzamiento interpoblacional y semilla $S_{0:1}$ de las mismas plantas. La semilla de familias de hermanos completos se utilizará para las evaluaciones del carácter bajo selección y la semilla $S_{0:1}$ se utilizará como unidad de recombinación en caso de que las familias sean seleccionadas.

A diferencia de la selección recurrente recíproca de medios hermanos que requería dos grupos de medios hermanos interpoblaionales para evaluar ($MH_A \times B$ y $MH_B \times A$), aquí solo es necesario evaluar un único grupo de familias interpoblaionales ($HC_{A \times B}$). El esquema de la selección recurrente recíproca de hermanos completos se presenta en la Figura 10.12.

En este esquema de selección la aplicación de los procedimientos de Eberhart y de Vencovsky para predecir la respuesta dan respectivamente las siguientes expresiones:

$$R = i(c')(rV_A)/(V_p)^{0.5} = \\ = i(1)(1/2)^{A \times B} V_A / \{ (1/2)^{A \times B} V_A + (1/4)^{A \times B} V_D + [(1/2)^{A \times B} V_{A \times E} + (1/4)^{A \times B} V_{D \times E}] / a + V_{EE} / ar \}^{0.5}$$

y

$$R = i\theta V_A / (V_p)^{0.5} = \\ = i [(1/2)(1)(1/2)(1/2) + (1/2)(1)(1/2)(1/2) + (1/2)(1)(1/2)(1/2) + (1/2)(1)(1/2)(1/2)]^{A \times B} V_A / \\ / [(1/2)^{A \times B} V_A + (1/4)^{A \times B} V_D + (1/2)^{A \times B} V_{A \times E} / a + (1/4)^{A \times B} V_{D \times E} / a + V_{EE} / ar]^{0.5} = \\ = i [4(1/8)]^{A \times B} V_A / [(1/2)^{A \times B} V_A + (1/4)^{A \times B} V_D + (1/2)^{A \times B} V_{A \times E} / a + (1/4)^{A \times B} V_{D \times E} / a + V_{EE} / ar]^{0.5} / \\ / a + (1/4)^{A \times B} V_{D \times E} / a + V_{EE} / ar]^{0.5} \\ = i [(1/2)]^{A \times B} V_A / [(1/2)_{A \times B} V_A + (1/4)_{A \times B} V_D + (1/2)_{A \times B} V_{A \times E} / a + (1/4)_{A \times B} V_{D \times E} / a + V_{EE} / ar]^{0.5}$$

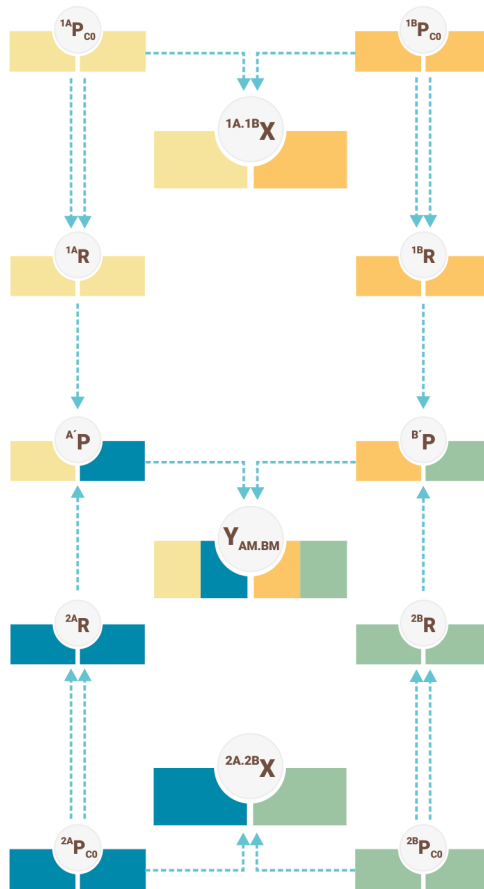


Figura 10.12. Esquema de la selección recurrente recíproca de hermanos completos (ver texto).

A diferencia de otros esquemas de selección recurrente recíproca, en el caso de la de hermanos completos vemos que existe un sendero que conecta, formando un ciclo completo, a un individuo Y del cruzamiento inter poblacional con los individuos de las dos poblaciones originales, A y B. Esta particularidad, permite capitalizar en la selección no solo los efectos aditivos, sino también parcialmente los de dominancia (Ehdaie y Cress, 1973).

En caso de que existiera epistasis contribuyendo al comportamiento de las familias inter poblacionales, la selección de los dos progenitores de cruzamientos destacados puede permitir que tales interacciones tengan oportunidad creciente de repetirse en subsiguientes ciclos, aportando a la respuesta a la selección (Jones *et al.*, 1971).

Uno de los atractivos de la selección recurrente recíproca de hermanos completos es la posibilidad de acoplarla con el desarrollo de los llamados

híbridos crípticos (Hallauer, 1967, 1973; Lonquist y Williams, 1967), en el caso de maíz. Los pares de familias $S_{0:1}$ correspondientes a las familias de hermanos completos interpoblacionales seleccionadas se siembran apareadas en las siguientes estaciones de cultivo con la finalidad de efectuar nuevos cruzamientos de hermanos completos interpoblacionales y nuevas autofecundaciones. Los cruzamientos $^A S_{1:2} \times ^B S_{1:2}$ entre plantas individuales dentro de ambas familias se evaluarán en experimentos al mismo tiempo que las semillas de autofecundación, $^A S_{1:2}$ y $^B S_{1:2}$ que resulten ser progenitores de cruza $^A S_{1:2} \times ^B S_{1:2}$ que resultarán seleccionadas se sembrarán para iniciar otro ciclo de cruzamientos y autofecundaciones ($^A S_{2:3} \times ^B S_{2:3}$, $^A S_{2:3} \times ^B S_{2:3}$) y así sucesivamente hasta obtener un híbrido promisorio y al mismo tiempo sus líneas progenitoras.

La selección recurrente recíproca de hermanos completos ha demostrado ser muy eficiente en maíz, pero también puede adoptarse en otros cultivos, como girasol. Márquez Sánchez (1982) presentó modificaciones a este procedimiento de selección en el caso de poblaciones de maíz que poseen típicamente una única espiga por planta.

Con algunas variantes, el método de selección recurrente recíproca de hermanos completos también puede utilizarse aunque no exista prolificidad en maíz. En lugar de efectuar cruzamientos $^A S_0 \times ^B S_0$, se siembran apareadas familias $^A S_{0:1}$ y $^B S_{0:1}$ y se realizan cruzamientos de varias plantas de A por varias plantas de B ($^A S_{0:1} \times ^B S_{0:1}$), al mismo tiempo que se autofecundan plantas de A y de B para generar familias $^A S_{0:2}$ y $^B S_{0:2}$.

Cada uno de los métodos de selección descritos admiten variantes que permiten adaptarlos a los objetivos de los programas de mejoramiento.

Si consideramos que las respuestas en el comportamiento de cada una de las poblaciones sometidas a selección recurrente recíproca son R_{AA} y R_{BB} (respuestas indirectas), la ganancia o respuesta directa esperada en el cruzamiento interpoblacional A x B (R_{AB}) sería:

$$R_{AB} = \frac{1}{2} (R_{AA} + R_{BB}) + R_h,$$

donde el término R_h es la respuesta o incremento esperado en la heterosis. Miranda Filho (1982) propuso dos diseños genéticos para poder estimar la respuesta en la heterosis.

10.6. Selección recurrente, número efectivo y endogamia

Las fórmulas de respuesta a la selección que se presentaron más arriba asumen que la población que se está mejorando posee tamaño infinito. En la práctica se trabaja sobre una muestra finita de la población, por lo que debería incorporarse en la fórmula de predicción de la respuesta la dependencia del tamaño efectivo poblacional. Es conocido que pequeños tamaños efectivos poblacionales pueden provocar, por la simple aleatoriedad de un muestreo limitado, la pérdida y la fijación de alelos generando situaciones de deriva genética y endogamia. Consecuentemente es posible encontrar, en caracteres que exhiben dominancia, disminuciones en la media poblacional por depresión por endocría. Recordemos la expresión de respuesta a la selección propuesta por Souza Jr. *et al.* (2000) que considera los efectos del tamaño finito poblacional:

$$R = [ic'(\sigma_A^2 + \Delta F.D_1) / \sigma_p] - (\Delta F)(ID),$$

donde:

ΔF es el incremento del coeficiente de endocría.

ID es la depresión por endocría.

D_1 es la covariancia entre los efectos aditivos y los de homocigotas dominantes.

Las predicciones de respuesta a la selección cuando se corrige por tamaño poblacional resultan inferiores a las predicciones realizadas con la expresión para tamaño infinito. En el caso de maíz, la comparación de respuestas esperables para distintos números efectivos indica que dicho número no debería ser menor de $N_e = 20$ para no "sacrificar" respuesta por deriva genética.

Obsérvese que en la etapa de recombinación de la selección recurrente resulta importante hacerlo de manera tal que cada una de las progenies (o individuos) seleccionados contribuyan con igual cantidad de semillas a la población mejorada. Ello permite ampliar el tamaño efectivo poblacional (ver sección correspondiente) y disminuir las probabilidades de que la endogamia actúe en dirección opuesta a la deseada por selección. De cualquier manera, tratándose de poblaciones relativamente pequeñas como las que se someten a selección recurrente, el incremento en el coeficiente de endocría por ciclo de selección será:

$$F = (1/2N_e)$$

Por lo tanto luego de t ciclos de selección, asumiendo la utilización del mismo tamaño efectivo en cada ciclo, el coeficiente de endogamia será:

$$F_t = [(1/2N_e)] + [1 - (1/2N_e)]F_{t-1}$$

Marquez-Sánchez (2009) dedujo la endogamia que generan diversos esquemas de selección recurrente en función del número de familias evaluadas y la intensidad de selección que se aplique y asumiendo que la recombinación de las familias seleccionadas se hace mediante cruzamiento dialélico y ajustando por la variancia del número de descendientes por familia.

En el caso de selección recurrente de medios hermanos, luego de t ciclos de selección, la endogamia esperada será:

$$F_{(MH)t} = [1/8(nm - 1)][(m - 1) + 8m(n - 1)F_{t-1} + 6(m - 1)F_{t-2} + (m - 1)F_{t-3}],$$

donde:

n es el número de familias que seleccionadas,

m es el número de plantas por familia seleccionada.

Para el caso de la selección de hermanos completos:

$$F_{(HC)t} = [1/4(nm-1)][(m-1) + 4m(n-1)F_{t-1} + 2(m-1)F_{t-2} + (m-1)F_{t-3}],$$

y para el caso de selección recurrente de familias endocriadas S_1 :

$$F_{(S1)t} = [1/2 (nm - 1)][(m - 1) + 2m(n - 1)F_{t-1} + (m - 1)F_{t-2}]$$

Como se indica más arriba, el tamaño efectivo poblacional es función del número de individuos que contribuyen a formar la generación siguiente y no necesariamente coincide con el número real de individuos. En el caso de la selección direccional por truncamiento es relativamente fácil determinar aproximadamente el número efectivo a partir del número y tipo de familias o individuos seleccionados (R) que se recombinarán (unidades de recombinación) para formar la población mejorada. De manera aproximada, Vencovsky (1980) indica que los tamaños efectivos poblacionales cuando se selecciona en forma recurrente de familias de hermanos completos y medios hermanos pueden calcularse a partir del número de familias seleccionadas (S):

$$N_{eSRHC} \approx 2S$$

y

$$N_{eSRMH} \approx 4S$$

Vencovsky (1978) sostiene que en la selección recurrente pueden existir una o más etapas en las cuales existe muestreo de gametas, y por lo tanto cada una de ellas sumará su efecto sobre el tamaño efectivo poblacional en cada ciclo selectivo. Este autor definió al tamaño efectivo poblacional basado en la variancia de las frecuencias alélicas debidas al efecto de la deriva genética (N_{ev} , en adelante N_e). Cuando se practica selección recurrente basada en algún tipo de familias, el muestreo de gametas ocurre en primer lugar cuando se seleccionan los individuos, y en segundo término cuando se muestrean gametas de las familias seleccionadas en la etapa de recombinación (a menos que las familias fueran de doble haploides duplicados)

Vencovsky (1978) demostró que:

$$N_e \approx N_{e1} \cdot N_{e2} / (N_{e1} + N_{e2}),$$

donde N_{ei} es el número efectivo debido al muestreo de gametas en la etapa i -ésima y que:

$$N_e \approx N_{eh} / 2,$$

donde N_{eh} es la media armónica de los tamaños efectivos en cada etapa.

Asumiendo que en una población monoica existe un número finito de N individuos en la cual se practica selección, un número S de ellos contribuirá con gametas masculinas y femeninas a la siguiente generación. Si no existe un control parental perfecto ($c=1$), los restantes individuos, $R=(N-S)$, también contribuirán con gametas masculinas, es decir, que solo se selecciona en uno de los sexos, o $c=0,5$. Si el control parental es $c=1$, entonces $R=0$. La proporción de individuos que contribuirán con gametas a la siguiente generación serán:

Cuando $c = 0,5$:

$$u = S/N \text{ para las gametas femeninas y}$$

$$v = (S+R)/N = M/N \text{ para las gametas masculinas.}$$

Cuando $c=1$:

$$u = S/N \text{ para las gametas femeninas y}$$

$$v = (S+0)/N = M/N = S/N \text{ para las gametas masculinas y por lo tanto}$$

$$u=v$$

En este último caso la selección se aplica con igual intensidad en los dos sexos.

Basado en estos conceptos, Vencovsky (1978) dedujo los tamaños efectivos poblacionales para diferentes métodos de selección recurrente:

Selección recurrente fenotípica ($c = 0,5$):

$$N_e = 4uN/(3u + 1) \text{ o}$$

$$N_e = 4uN/(2u + 1),$$

Para la aplicación de estas fórmulas debe considerarse si se controla o no el número de individuos (semillas) con que cada individuo hembra contribuye a la población mejorada, respectivamente. Nótese que cuando se anula la variancia en el número de gametas contribuida a la siguiente generación, el tamaño efectivo poblacional es algo mayor.

Selección recurrente fenotípica sin recombinación o masal (en especies autóгамas)

$$N_e \approx uN/2(1 - u) \text{ si } N \text{ es suficientemente grande para que } N/(N-1) \approx 1$$

Selección recurrente basada en algún tipo de familias y recombinación de individuos S_1 .

Se considera en este caso que de cada una de las S familias seleccionadas se muestrean P individuos S_1 para efectuar la recombinación y se vuelven a tomar N individuos para iniciar el siguiente ciclo de selección. Ello implica que existen dos etapas de muestreo que condicionan el tamaño efectivo:

$$N_{e1} = S/(1 - u + 1/2P) \text{ y } N_{e2} = N = S/u$$

$$N_e \approx N_{eh}/2 = uN/(1 + 1/2P)$$

Selección recurrente basada en algún tipo de familias y recombinación de individuos S_1 tomando igual cantidad de individuos de cada progenie en la etapa de recombinación:

$$N_e \approx N_e h/2 = S/(1 + 1/P)$$

10.7. Selección recurrente genómica

En la selección recurrente tradicional el criterio de selección es la magnitud del valor de mejora de los diferentes individuos estimados a partir de datos fenotípicos colectados sobre la progenie de esos mismos individuos. En la variante de la selección recurrente genómica, en cambio, se utiliza como criterio de selección a las predicciones de los valores de mejora genómicos; esto es posible a partir del modelado de datos fenotípicos y la

caracterización genotípica utilizando una alta densidad de marcadores moleculares aleatoriamente distribuidos en el genoma. Las estimaciones del valor de mejora genómico (EVMG) pueden ser empleadas para la selección artificial, de acuerdo al siguiente modelo (Iwata y Jannink, 2011):

$$y_i = \mu + \sum \beta_j x_{ij},$$

donde:

y_i es el EVMG del individuo i -ésimo,

β_j es el vector de estimaciones de los efectos de los j -ésimos marcadores y

x_{ij} es el genotipo del i -ésimo individuo para cada j -ésimo marcador.

Si se utilizan las EVMG con fines de selección, deberemos preocuparnos por lograr que estas estimaciones se acerquen lo más posible a los verdaderos valores de mejora (VVM) de los individuos de la población bajo selección, es decir, que el nivel de precisión de las EVMG sea elevado. Para ello, debería procurarse que la mayoría de los QTLs que controlan el carácter de interés estén en desequilibrio de ligamiento con al menos uno de los marcadores moleculares. En tales condiciones, la matriz Q tenderá a ser G y proporcionalmente R irá tendiendo a cero y disminuyendo en importancia (ver definición de las matrices Q , G y R en el Capítulo 3). Para lograrlo hace falta utilizar un número de marcadores p , que deberá ser tanto mayor cuanto más complejo sea el carácter en términos del número de QTLs que operan en su control.

Para los fines de la selección nos interesa definir los siguientes términos:

r_{Q^*Q} es la correlación entre Q^* (o EVMG) y Q (o efectos genéticos captados por los marcadores moleculares en desequilibrio de ligamiento con los QTLs que controlan el carácter bajo selección). Cuanto mayor sea r_{Q^*Q} , mayor será la precisión de Q^* como estimador de Q .

$r_{GQ^*} = r_{MG}$ es la correlación entre G (o VVM) y Q^* (o EVMG). Esta correlación mide la precisión del EVMG como estimador del VVM.

q^2 es la variancia genética o aditiva explicada por Q .

La relación entre las dos correlaciones r_{MG} y r_{Q^*Q} es (Dekkers, 2007):

$$r_{MG} = q r_{Q^*Q}$$

Cuando se utilizan modelos alternativos para aplicar en programas de selección genómica es importante distinguir entre la habilidad predictiva del

modelo y la precisión del modelo. La primera será mayor cuanto más grande sea la correlación entre los valores observados y los valores que predice el modelo. La precisión (*accuracy*) del modelo, en cambio, mide la correlación entre los valores predichos por el modelo y los verdaderos valores de mejora (r_{MG}).

Simbolizaremos la habilidad predictiva del modelo como $r_{(y,Q^*)}$ y la precisión del modelo como $r_{(Q^*G)}$, equivalente a r_{MG} .

La habilidad predictiva y la precisión del modelo son diferentes, pero se relacionan una con la otra por la expresión (Dekkers, 2007; Albrecht *et al.*, 2011):

$$r_{Q^*G} = r_{MG} = r_{y,Q^*} / h,$$

donde h es la raíz cuadrada de la heredabilidad en sentido estricto de los valores observados (fenotípicos).

La precisión del modelo será mayor cuanto menor la heredabilidad, mientras que la habilidad predictiva será mayor cuanto mayor sea la heredabilidad. Ello muestra la relación de compromiso entre la precisión y la habilidad predictiva del modelo de estimación. Al fitomejorador le interesará prioritariamente utilizar un modelo de alta precisión, pues esta característica está directamente ligada con la ganancia genética por selección.

Existe una amplia lista de métodos alternativos de estimación del valor de mejora genómico (EVMG). Cuando el número de marcadores cuyo efecto queremos estimar supera largamente al número de genotipos, se plantean dos dificultades: i) la falta de grados de libertad para hacer las estimaciones de los efectos de tantos marcadores y ii) la multicolinealidad o correlación entre efectos de marcadores, que sobreajustan el modelo y reducen su capacidad predictiva. Para solucionarlo se emplean modelos estadísticos apropiados que seleccionan variables, encogen los efectos o ambas cosas a la vez (De los Campos *et al.*, 2013; Lin *et al.*, 2014). Los modelos de estimación pueden ser paramétricos o no paramétricos o semiparamétricos.

Entre los primeros están:

- *Ridge Regression* - BLUP.
- BLUP genómico.
- Poligénico.
- Bayesianos (A, B, LASSO, Cp, R).
- Regresión Parcial de Mínimos Cuadrados.
- Componentes principales.



Entre los no paramétricos y semiparamétricos:

- RKHS.
- Máquinas de aprendizaje: SVR, RF, Redes Neuronales.

Las características principales de estos métodos pueden consultarse en De los Campos *et al.* (2013); Lin *et al.* (2014) y Jannink *et al.* (2010).

Debido a que hay una relación lineal y directa entre la precisión del modelo (r_{MG}) y la ganancia o respuesta a la selección genómica, resulta apropiado describir el ciclo de selección genómica y luego analizar cuales son los factores que afectan la precisión del modelo.

La selección genómica o selección recurrente genómica es en esencia un proceso cíclico semejante a la selección recurrente convencional, que alterna diferentes métodos de evaluación de genotipos, tanto fenotípicos como moleculares. La selección se basa en las estimaciones de los valores de mejora genómicos (EGVM) de las unidades de selección, pero para ello necesitamos contar con un método de estimación de los efectos de los marcadores a partir de datos fenotípicos y moleculares. Para proceder a la selección genómica debemos disponer de dos poblaciones: una población de entrenamiento (o de referencia) y una población de cría o candidata.

La población de entrenamiento requiere de ambos tipos de datos, fenotípicos y de marcadores de ADN en alta densidad, correspondientes a los individuos que la componen. A partir de ellos se corren algunos de los modelos indicados más arriba y se escogerá uno de ellos para obtener los EVMG (matriz Q^*).

La población de cría es la población que interesa seleccionar. Normalmente sus individuos solo serán evaluados genotípicamente y su mérito genético o valor reproductivo estimado por los EVMG (Q^*). Los individuos seleccionados en forma direccional por truncamiento por superar un valor mínimo de EVMGs serán recombinados para generar la población mejorada (C_1). Sobre esta población pueden volver a seleccionarse individuos con base en los EVMG por uno o dos ciclos más, hasta que se hace necesario volver a modelar los efectos de los marcadores y reiniciar el proceso (Heffner *et al.*, 2009) (Figura 10.13).

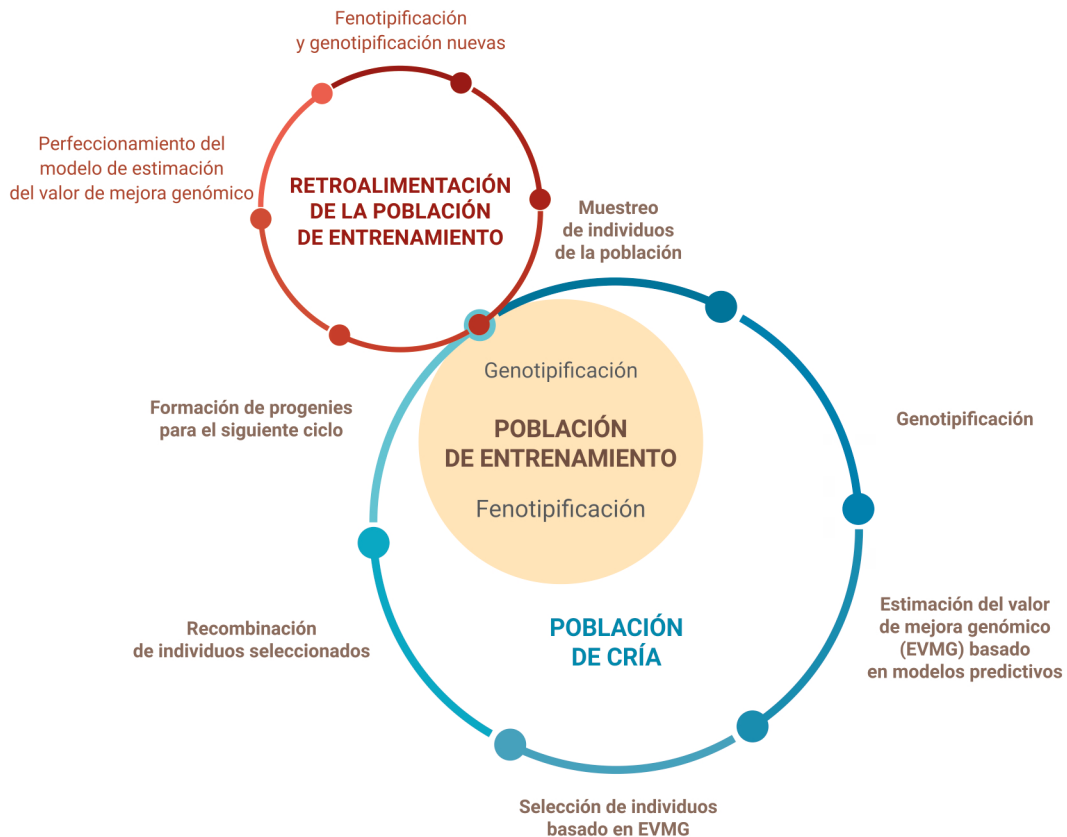


Figura 10.13. Esquema general de la selección recurrente genómica.

Anteriormente señalamos que resulta crítico lograr captar en Q^* la mayor parte de la variabilidad aditiva en la población y ello depende del desequilibrio de ligamiento entre marcadores y QTLs. La propia selección afecta el desequilibrio de ligamiento, de manera que en la medida que se vayan sucediendo ciclos de selección y recombinación en la población, la precisión del modelo para estimar los EVMG se reduce concomitantemente. Por ello resulta importante ajustar el modelo periódicamente, luego de dos a cuatro ciclos de selección basada en EVMG, según las características de la selección (hábito reproductivo de la especie, tamaño efectivo poblacional, etc.).

Los factores que afectan directamente la precisión del modelo de estimación de los EVMGs son el desequilibrio de ligamiento y las relaciones de parentesco. Los factores que indirectamente afectan a la precisión lo hacen por su efecto directo sobre los dos factores anteriores. Entre ellos podemos mencionar el tamaño efectivo, el tamaño y las características de

la población de entrenamiento, el método de estimación empleado, la complejidad del carácter bajo selección (número de QTLs y heredabilidad) y las interacciones entre estos factores.

10.7.1. Características de la población de entrenamiento

Las características apropiadas que debe reunir una población de entrenamiento son que la variancia explicada por los marcadores sea lo mayor posible, que guarde relación con la diversidad genética de la población de cría y del programa de mejoramiento y que permita reducir la multicolinealidad entre los marcadores. Reducir la colinearidad implica aumentar la precisión en la estimación de los EVMG). Para ello resulta conveniente fenotipificar individuos de progenies que han tenido un mayor número de eventos de recombinación. Desde el punto de vista genético resulta necesario que haya cierto grado de relación con la población de cría y que el tamaño de la población sea grande. Mayores tamaños de la población de entrenamiento mejoran la precisión y disminuyen la proporción esperada de alelos monomórficos. Por otra parte conviene que la población de entrenamiento esté en desequilibrio de ligamiento y que la densidad de marcadores sea elevada, de forma tal que aumente la probabilidad de ligamiento entre marcadores y QTLs y que cada QTL esté en desequilibrio de ligamiento con al menos un marcador.

Asumiendo que se utilizan marcadores (N_M) en número suficiente para saturar el genoma, la precisión del modelo sigue la siguiente expresión (Combs y Bernardo, 2013; Lin *et al.*, 2014):

$$r_{Q^*G} = [Nh^2/(Nh^2 + M_e)]^{1/2} \quad [10.2]$$

siendo:

N el número de individuos genotipificados y fenotipificados en la población de entrenamiento.

h^2 la heredabilidad del carácter.

M_e el número efectivo de segmentos cromosómicos independientes.

Esté último término (M_e) puede estimarse a partir del tamaño efectivo de la población y la longitud total del genoma (Lin *et al.*, 2014):

$$M_e = 2N_e L,$$

donde

N_e es el tamaño efectivo de la población de entrenamiento y

L es la longitud del genoma (medido en Morgans).

Reemplazando en [10.2]

$$r_{Q*G} = [Nh^2 / (Nh^2 + 2N_eL)]^{1/2}$$

Cuanto más emparentados estén entre sí los individuos de la población de entrenamiento, mayor será la longitud de los segmentos cromosómicos independientes y por lo tanto menor será el número efectivo de segmentos cromosómicos independientes (menor M_e), por lo que pueden estimarse M_e para distintos tipos de familias. Otra manera de estimar M_e es hacerlo empíricamente a partir del grado de desequilibrio de ligamiento promedio (r^2).

Cuanto mayor sea el número efectivo de segmentos cromosómicos independientes, la precisión tenderá a ser menor (ya que aumentará el requerimiento en número de marcadores que deben estar en ligamiento con los segmentos).

Cuando el número de individuos en la población de entrenamiento es mayor o igual que el número de segmentos cromosómicos independientes (M_e), debe emplearse una fórmula corregida, de modo tal que:


$${}^c r_{Q*G} = r_{Q*G} + r_{Q*G}^4 M_e / 2N \text{ (Daetwyler et al., 2008; Hayes et al., 2013).}$$

10.7.2. Complejidad del carácter bajo selección y precisión del modelo

Cuanto más sencilla sea la herencia del carácter (mayor heredabilidad y menor el número de QTLs) la precisión del modelo aumenta (Resende et al., 2011). Sin embargo aun a baja heredabilidad, con la selección genómica puede obtenerse mayor respuesta que con la selección convencional (Lin et al., 2014). Cuando la interacción genotipo x ambiente resulta importante, ello compromete la precisión del modelo (Heslot et al., 2013).

10.7.3. Efecto del tamaño efectivo sobre la precisión del modelo

La disminución del tamaño efectivo de la población de entrenamiento implica una mayor tasa de endocria y probabilidad de deriva genética. Estos factores incrementan el desequilibrio de ligamiento y consecuentemente mejoran la predicción del efecto de los QTLs a partir de los marcadores.



Por otra parte la disminución del tamaño efectivo provoca una disminución del número de segmentos cromosómicos independientes (M_e), a condición de que el número de individuos fenotipados y genotipados de la población de entrenamiento (N) y la heredabilidad se mantengan constantes. Ambas situaciones, estas son la mejora en la predicción de los efectos y la disminución del número de segmentos cromosómicos independientes, conducen a un incremento en la precisión del modelo.

El hábito reproductivo de una especie también afecta el tamaño efectivo poblacional. Las poblaciones de especies alógamas como por ejemplo el maíz, el eucalipto o el *ryegrass* tienden a tener mayor tamaño efectivo poblacional que las de especies autógamas como la cebada o el trigo. Desde el punto de vista del número de segmentos cromosómicos independientes, es de esperar que este sea mayor en las especies alógamas que en las autógamas y en consecuencia podría esperarse que la precisión del modelo fuera *a priori* superior en las especies autógamas. Sin embargo, en los esquemas de selección artificial pueden utilizarse diferentes tipos de familias, pudiendo ocurrir variaciones dentro de la misma especie y situaciones en las que determinadas progenies de especies alógamas permiten lograr menor número de segmentos cromosómicos independientes y mayor precisión (Lin *et al.*, 2014).

10.7.4. Número y tipo de marcadores necesarios y precisión del modelo

Meuwissen (2009) describe la siguiente relación entre el número de marcadores necesarios para alcanzar una buena precisión del modelo en función del tamaño efectivo poblacional (N_e) y la longitud del genoma (L):

$$N_m \geq 10LN_e$$

Como se indica más arriba, cuanto mayor sea el tamaño efectivo, mayor será el número esperado de segmentos cromosómicos independientes que deberán resultar asociados (o “marcados”) con al menos un marcador.

Cuando se realiza selección dentro de familias, sus individuos emparentados tienden a compartir largos segmentos cromosómicos y solo ese número de segmentos requerirá ser “marcado”. En este caso de selección intrafamiliar (Hayes *et al.*, 2009) o dentro de familias el número de marcadores será

$$N_m < 10LN_e$$

El número de marcadores necesarios también dependen de su tipo o naturaleza. Este será menor cuanto mayor sea el número de alelos que permiten

detectar. Por ejemplo, si se emplean microsatélites o haplotipos de múltiples SNPs, probablemente sea necesaria menor cantidad que si se utilizan marcadores SNPs tomados individualmente (Solberg *et al.*, 2008, Lin *et al.*, 2014).

10.7.5. Densidad de marcadores y precisión del modelo

Para una misma cantidad de marcadores y con distinto tamaño del genoma, el espaciamiento entre marcadores será diferente. Cuanto más extenso es el genoma, más distanciados estarán esos marcadores, aumentará la probabilidad de recombinación entre ellos, se reducirá el desequilibrio de ligamiento y se afectará negativamente la precisión del modelo (Muir, 2007).

Para establecer la densidad de marcadores apropiada debe tenerse en cuenta la relación entre la tasa de decaimiento del desequilibrio de ligamiento y el aumento de la distancia genética (Heffner *et al.*, 2009). No obstante ello difiere considerablemente entre genes (Remington *et al.*, 2001).

10.7.6. Heredabilidad del carácter y precisión

Recordemos la expresión:

$$r_{Q^*G} = [Nh^2 / (Nh^2 + M_e)]^{1/2}$$

No obstante ello, para un carácter determinado, Combs y Bernardo (2013) concluyen que lo que determina la precisión no es el tamaño de la población de entrenamiento (N) ni la heredabilidad del carácter (h^2) tomados por sí solos, sino el producto de ambos factores.

10.7.7. Desequilibrio de ligamiento y precisión del modelo

La tasa y patrón de decaimiento del desequilibrio de ligamiento depende de la especie, población genética, región genómica, historia evolutiva, sistema de apareamiento, tamaño poblacional, estructura poblacional, tasa de recombinación y efectos de la selección (Gaut y Long, 2003, Heffner *et al.*, 2009).


10.7.8. Respuesta a la selección genómica

Recordemos la expresión de respuesta a la selección fenotípica:

$${}^P R = ih \sigma_A = ih^2 \sigma_P$$

En el caso de la selección genómica:

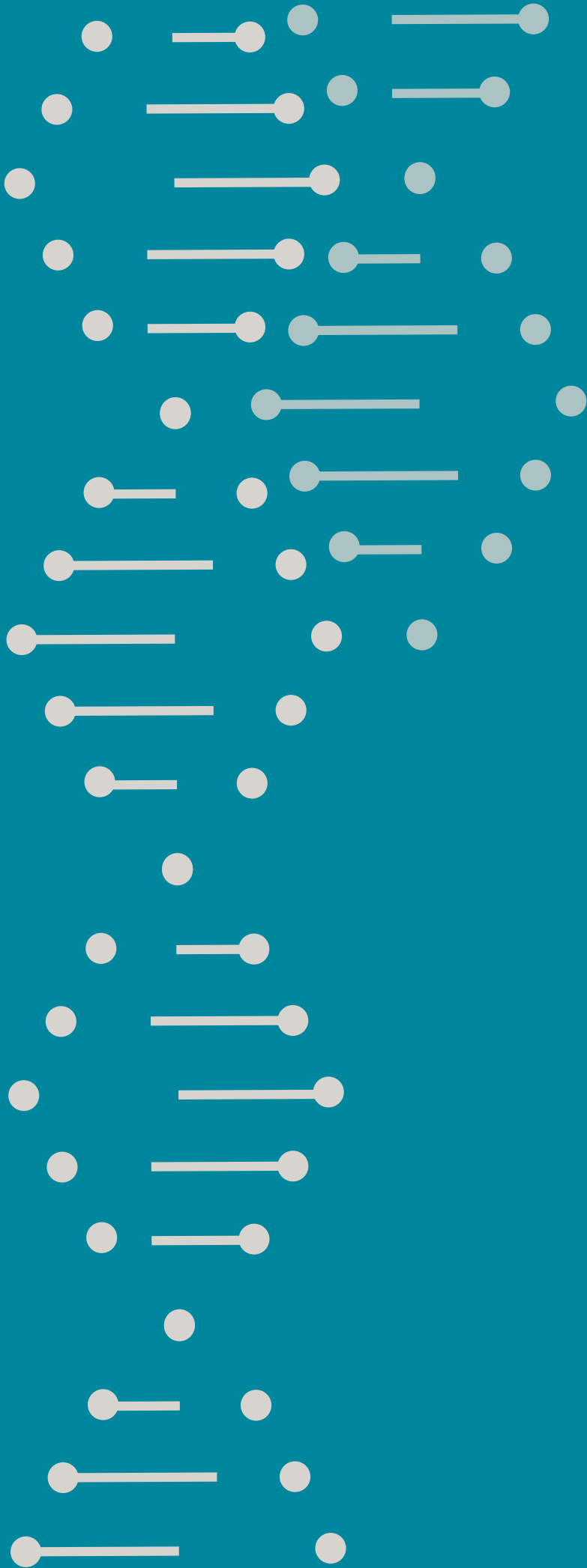
$${}^G R = ir_{Q^*G} \sigma_A$$



donde el término de raíz cuadrada de la heredabilidad es reemplazado por la correlación entre los EVMG y los VVM (precisión de selección o *selection accuracy*).

$${}^G R = i r_{(Q^*, G)} \sigma_A = i \{ [N h^2 / (N h^2 + M_e)]^{1/2} \} \sigma_A$$

donde σ_A es el desvío estándar de los verdaderos valores de mejora.





Correlación entre caracteres

Los caracteres cuantitativos pueden presentar asociaciones o correlaciones estadísticas entre sus valores fenotípicos. En términos estadísticos, la correlación fenotípica entre un par de caracteres, X e Y, (r_{XY}^P) es la relación entre su covarianza fenotípica y el producto de sus desvíos estándares fenotípicos. Análogamente pueden definirse las correlaciones genotípicas (r_{XY}^G) y ambientales (r_{XY}^E). De tal manera:

$$r_{XY}^P = \frac{P\sigma_{XY}}{P\sigma_X P\sigma_Y}; P\sigma_{XY} = r_{XY}^P P\sigma_X P\sigma_Y \quad [11.1]$$

$$r_{XY}^G = \frac{G\sigma_{XY}}{G\sigma_X G\sigma_Y}; G\sigma_{XY} = r_{XY}^G P\sigma_X P\sigma_Y$$

$$r_{XY}^E = \frac{E\sigma_{XY}}{E\sigma_X E\sigma_Y}; E\sigma_{XY} = r_{XY}^E P\sigma_X P\sigma_Y$$

donde:

${}^P\sigma_{XY}$, ${}^G\sigma_{XY}$, ${}^E\sigma_{XY}$ son, respectivamente, la covarianza fenotípica, genotípica y ambiental entre X e Y,

${}^P\sigma_X$, ${}^G\sigma_X$, ${}^E\sigma_X$ son los desvíos estándares fenotípico, genotípico y ambiental del carácter X,

${}^P\sigma_Y$, ${}^G\sigma_Y$, ${}^E\sigma_Y$ son los desvíos estándares fenotípico, genotípico y ambiental del carácter Y.

Los valores fenotípicos (observados) para las variables X e Y en el i-ésimo individuo responden a la media general, al valor genotípico y al desvío ambiental respectivo:

$$X_i = \mu_X + G_{Xi} + e_{Xi}$$

$$Y_i = \mu_Y + G_{Yi} + e_{Yi}$$

La covarianza de los valores fenotípicos será:

$${}^PCov_{(Xi, Yi)} = {}^P\sigma_{XY} = \varepsilon (G_{Xi} + e_{Xi}, G_{Yi} + e_{Yi}) = {}^G\sigma_{xy} + {}^E\sigma_{xy}, \quad [11.2]$$

porque los productos cruzados ($e_{Xi}e_{Yi}$) son iguales a cero puesto que e_{Xi} y e_{Yi} son independientes uno del otro.

Considerando [11.1] y [11.2] resulta que:

$${}^Pr_{XY} {}^P\sigma_X {}^P\sigma_Y = {}^Gr_{XY} {}^P\sigma_X {}^P\sigma_Y + {}^Er_{XY} {}^P\sigma_X {}^P\sigma_Y;$$

Recordemos que

$${}^G\sigma_X = H_x {}^P\sigma_X$$

$${}^G\sigma_Y = H_y {}^P\sigma_Y$$

$${}^E\sigma_X = e_x {}^P\sigma_X$$

$${}^E\sigma_Y = e_y {}^P\sigma_Y$$

donde H_x y H_y son las raíces cuadradas de las heredabilidades en sentido amplio. Consecuentemente:

$${}^Pr_{XY} {}^P\sigma_X {}^P\sigma_Y = ({}^Gr_{XY} H_x {}^P\sigma_X H_y {}^P\sigma_Y) + ({}^Er_{XY} e_x {}^P\sigma_X e_y {}^P\sigma_Y)$$

Multiplicando ambos miembros de la igualdad por $1/{}^P\sigma_X {}^P\sigma_Y$ se puede reescribir la ecuación como:

$${}^Pr_{XY} = ({}^Gr_{XY} H_x H_y) + ({}^Er_{XY} e_x e_y)$$

Es decir, que la correlación fenotípica es la suma de las correlaciones genotípicas y ambientales ponderadas, respectivamente, por el producto de las raíces cuadradas de la heredabilidad de los dos caracteres y por el producto de la relación (e_x, e_y) entre los desvíos estándares ambientales y los desvíos estándares fenotípicos de los dos caracteres correlacionados.

La correlación genotípica puede descomponerse, a su vez, en correlación genética o entre los valores de mejora, correlación entre los desvíos de dominancia y correlación entre los efectos de interacción epistática. En razón de la complejidad del tratamiento de estas dos últimas, se las incorpora como componentes de la correlación ambiental. Por tal motivo, la correlación ambiental resulta tanto de los desvíos ambientales como de los componentes no aditivos de la correlación genotípica. Desde un punto de vista práctico, el componente que interesa tener en cuenta en programas de selección es la correlación genética o aditiva puesto que es la responsable de las respuestas indirectas (o correlacionadas) a la selección. Esencialmente podemos hacer una analogía con la variancia fijable y decir que la correlación genética es la que importa considerar en las diferentes estrategias de selección.

Para obtener una expresión que relacione la correlación fenotípica con la correlación genética (aditiva) y la ambiental podemos seguir un procedimiento de deducción similar al anterior:

$$\begin{aligned} X_i &= \mu_x + A_{Xi} + e_{Xi} \\ Y_i &= \mu_y + A_{Yi} + e_{Yi} \end{aligned}$$

En este caso, los efectos de dominancia y de epistasis, de ser distintos de cero, quedan incluidos en los componentes ambientales e_x y e_y .

La covariancia de los valores fenotípicos será:

$${}^P\text{COV}_{(X_i, Y_i)} = {}^P\sigma_{XY} = \varepsilon(A_{Xij} + e_{Xij}, A_{Yij} + e_{Yij}) = {}^A\sigma_{xy} + {}^E\sigma_{xy}, \quad [11.3]$$

Reemplazando términos y operando algebraicamente llegamos a la siguiente expresión:

$${}^P r_{XY} = h_x h_y {}^A r_{xy} + e_x e_y {}^E r_{xy}, \quad [11.4]$$

donde:

h_x y h_y son las raíces cuadradas de la heredabilidad en sentido estricto,

e_x y e_y son las raíces cuadradas de la relación entre la variancia ambiental (incluyendo eventualmente componentes de origen genético no aditivos) y la variancia fenotípica de cada uno de los caracteres.

En el modelo [11.4], la correlación fenotípica, cualquiera sea su magnitud y signo, resulta de dos causas concurrentes: la correlación genética o aditiva y la correlación ambiental. Si bien la magnitud de la correlación fenotípica responde a estas razones causales, no es simplemente la suma de ambas. Por el contrario, la correlación fenotípica es la suma de las correlaciones genética (o aditiva) y ambiental ponderadas, respectivamente, por el producto de las raíces cuadradas de la heredabilidad en sentido estricto de los dos caracteres y por el producto de la relación entre los desvíos estándares ambientales y los desvíos estándares fenotípicos de los dos caracteres correlacionados.

Por sus consecuencias sobre la selección, la correlación genética es la más importante, y puede deberse a dos fenómenos: la pleiotropía y el ligamiento genético. La pleiotropía es la asociación entre caracteres debida a que están total o parcialmente controlados por un grupo de genes en común. Por lo tanto, es una causa permanente de correlación genética. En cambio, el desequilibrio de ligamiento entre loci constituye una causa transitoria de correlación, ya que se mantendrá hasta que los eventos de recombinación genética entre loci en cada generación interrumpen o extingan esa asociación.

La ausencia de correlación fenotípica no implica necesariamente la inexistencia de correlación genética, debido a que una situación de correlación ambiental de signo opuesto puede contrarrestar la expresión de la primera.

La existencia de correlación genética entre caracteres se refleja en las respuestas a la selección. La selección para un carácter generará una respuesta indirecta en otro, siempre que estén genéticamente correlacionados (Fig. 11.1), y en el mismo sentido (signo) de la correlación entre ellos.

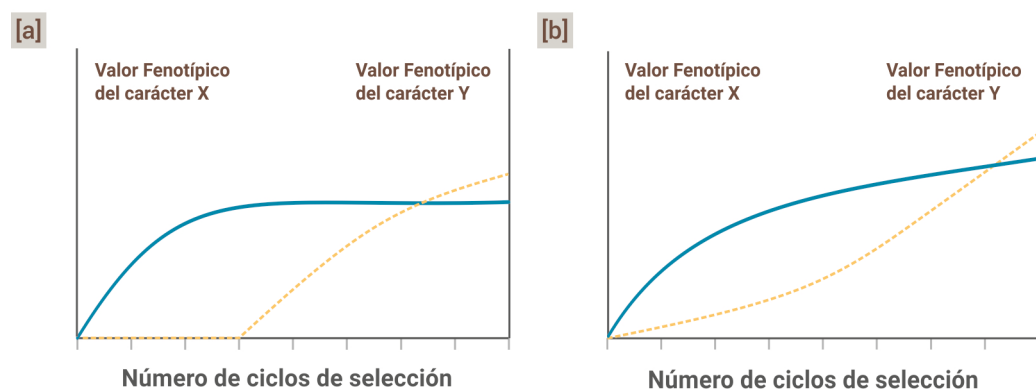


Figura 11.1. Efecto de la selección en tándem (ver el Capítulo 12) sobre dos caracteres, X e Y. [a]: X e Y no están correlacionados y por lo tanto no hay respuesta indirecta a la selección (la selección por X no provoca cambios en Y y viceversa). [b]: X e Y están positivamente correlacionados, de manera que la selección para X provoca aumentos en Y y viceversa.

Cuando se selecciona simultáneamente para dos o más caracteres genéticamente correlacionados, el sentido o la dirección de la selección para los caracteres involucrados puede generar cambios en la magnitud de la correlación y si se sostiene por un número suficiente de ciclos podrían darse modificaciones en el signo de la correlación (Fig. 11.2).

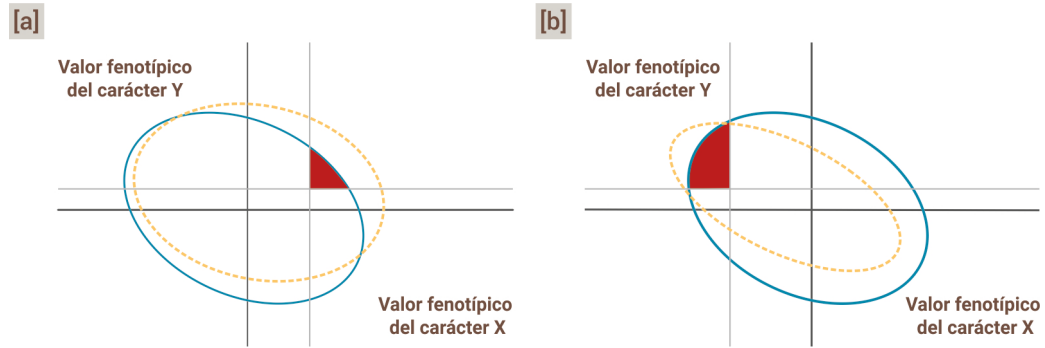


Figura 11.2.

Efecto de la selección simultánea para dos caracteres empleando niveles de cortes independientes (ver texto) sobre el valor fenotípico de dos caracteres, X e Y. [a]: Ambos caracteres están negativamente correlacionados y la selección es direccional hacia mayores valores fenotípicos en cada carácter. La fracción rayada indica los individuos seleccionados. La elipse punteada indica la probable distribución bivariada de los individuos luego de un número de ciclos de selección. [b] Ambos caracteres están negativamente correlacionados y la selección es direccional hacia menores y mayores valores fenotípicos en el carácter X e Y, respectivamente. La fracción rayada indica los individuos seleccionados. La elipse punteada indica la distribución probable de los individuos luego de un número de ciclos de selección.

Consideremos una estructura de asociación entre un carácter que se desea mejorar (D) y otros tres caracteres (A, B, C) correlacionados con D y algunos de ellos también correlacionados entre sí. Dependiendo de los signos de las correlaciones, las relaciones entre los caracteres en tal estructura pueden catalogarse de diferentes maneras desde el punto de vista de la resultante de la selección por A, B o por C con el objetivo final de mejorar D (Figura 11.3).

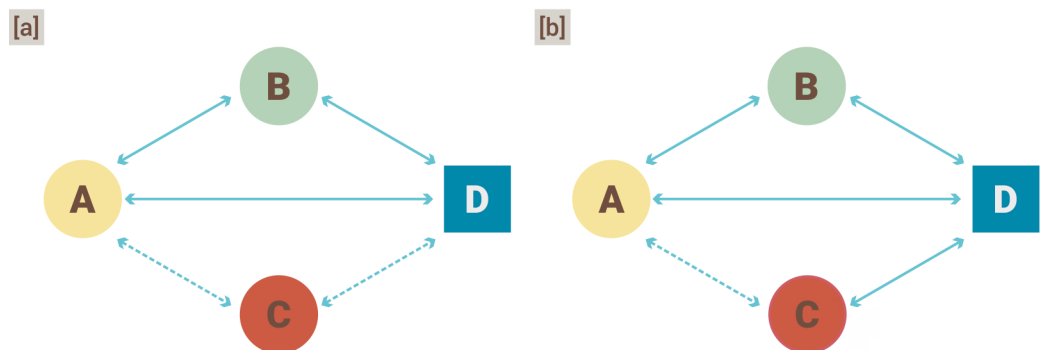


Figura 11.3.

Estructura de correlaciones genéticas entre tres caracteres, A, B y C con el carácter D, objetivo de mejoramiento. Los esquemas [a] y [b] difieren en el signo de las asociaciones. Las dobles flechas continuas y punteadas indican correlaciones positivas y negativas, respectivamente. Adaptado de Mariotti, 1986.

Los caracteres A, B y D, positivamente correlacionados entre sí, se dice que guardan una relación de complementación (Mariotti, 1986). Si se seleccionara por A, ello provocaría respuestas indirectas sobre D en la misma dirección que si la selección se practicara directamente por D. Al existir una correlación positiva de A con B, el efecto indirecto de la selección por A sobre B también resultará en una respuesta favorable en D, puesto que la correlación entre B y D es positiva. Aplicando el mismo razonamiento llegaríamos a la misma conclusión si el carácter seleccionado fuera B. Otro caso de complementación puede darse cuando se plantea la selección directa sobre A buscando una respuesta indirecta positiva sobre D. Si bien esto es posible por cuanto A y D están positivamente correlacionados, puede existir un efecto complementario por la correlación negativa, por un lado, entre A y C y por otro lado entre C y D [a].

Un escenario de incompatibilidad ocurriría si se intentara seleccionar por C para mejorar D. La resultante sería negativa, no solo por la correlación negativa entre C y D, sino también la que existe entre C y A, y la circunstancia de que A y D están positivamente relacionados (Fig. 11.3a).

Otras estructuras de asociación entre caracteres son las de compensación y equilibrio. Esencialmente son muy parecidas, y podríamos abarcarlos bajo el término de neutralidad. Sería el caso de la selección por A, para mejorar indirectamente a D. Como la correlación entre A y C es negativa, y entre C y D positiva, la respuesta por la vía AD puede ser contrarrestada, parcial o totalmente por lo que ocurra con la estructura de asociación que une A, C y D (Fig. 11.3b). La resultante final dependerá de la magnitud de las correlaciones, además de su signo.

11.1. Estimación de la correlación fenotípica

Una de las maneras de estimar la correlación fenotípica es mediante el cálculo del coeficiente de correlación intraclase aplicado a una variable clasificatoria (genotipos) para las dos variables aleatorias (caracteres) cuya asociación se quiere determinar. Supongamos que se evalúa un conjunto de familias de medios hermanos para dos caracteres, X e Y. Las esperanzas de los cuadrados medios para X y para Y y la de los productos medios XY serán como se indica en la Tabla 1.11.



Fuente de variación	ϵ (CM _x)	ϵ (CM _y)	ϵ (PM _{xy})
Entre Familias	$\sigma_{wx}^2 + k\sigma_{bx}^2$	$\sigma_{wy}^2 + k\sigma_{by}^2$	$\sigma_{wxwy} + k^*\sigma_{bxbx}$
Dentro de Familias	σ_{wx}^2	σ_{wy}^2	σ_{wxwy}

*k es el número individuos evaluados dentro de familias.

Tabla 11.1.

Partición de la suma de cuadrados de genotipos en componentes inter- e intrafamiliares para dos caracteres, X e Y, y esperanzas de los cuadrados medios del ANdeVA y de los productos medios del análisis de covarianza (ANCOVA). Los subíndices w y b hacen referencia, respectivamente, a los componentes de variancia dentro y entre familias.

La correlación fenotípica total de X con Y puede estimarse a partir de los componentes de variancia y covarianza obtenidos de la resolución de tres sistemas (CM_x, CM_y y PM_{xy}) de dos ecuaciones cada uno.

Así:

$${}^P r_{xy} = ({}^W \sigma_{xy} + {}^B \sigma_{xy}) / [({}^W \sigma_x^2 + {}^B \sigma_x^2)({}^W \sigma_y^2 + {}^B \sigma_y^2)]^{0,5}$$

y la correlación fenotípica dentro de familias de X con Y será:

$${}^W r_{xy} = ({}^W \sigma_{xy}) / [({}^W \sigma_x^2)({}^W \sigma_y^2)]^{0,5}$$

Otro procedimiento sencillo para calcular la correlación fenotípica consiste en calcular el coeficiente de correlación de Pearson sobre los datos fenotípicos, aunque la información que provee es menor que utilizando el ANdeVA y ANCOVA. Tengamos en cuenta que en este último caso podemos probar la significancia de las fuentes de variación **entre** y **dentro** de familias y mediante un experimento con repeticiones podremos estimar y controlar en alguna medida el error experimental.

11.2. Estimación de la correlación genética

Existen tres maneras básicas de estimar las correlaciones genéticas entre caracteres:

- A partir de las estimaciones de componentes de variancia y covariancias en análisis de familias.
- Empleando análisis padre-progenie para los caracteres X e Y.
- Efectuando un experimento de selección para un carácter X y evaluando las respuestas para los dos caracteres X e Y.

11.2.1. Estimación a partir de evaluación de familias

Suponiendo que se evalúan familias de medios hermanos, las esperanzas de los cuadrados medios del análisis de variancia y de los productos medios serán los que se indican en la Tabla 11.2.

Fuente de variación	ϵ (CM _x)	ϵ (CM _y)	ϵ (PM _{xy})
Entre familias de medios hermanos	$\sigma_{ex}^2 + r\sigma_{bx}^2 =$	$\sigma_{ey}^2 + r\sigma_{by}^2 =$	$\sigma_{exey} + r\sigma_{bxbx} =$
	$\sigma_{ex}^2 + r \frac{1}{4} \sigma_{Ax}^2$	$\sigma_{ey}^2 + r \frac{1}{4} \sigma_{Ay}^2$	$\sigma_{exey} + r \frac{1}{4} \sigma_{Ax Ay}$
Error experimental	σ_{ex}^2	σ_{ey}^2	σ_{exey}

Tabla 11.2.

Esperanzas de los cuadrados medios del ANdeVA y de los productos medios del análisis de covariancia (ANCOVA) para dos caracteres X e Y, evaluados en familias de medios hermanos en experimentos con r repeticiones.

Recordemos que las variancias y covariancias entre medios hermanos expresan 1/4 de la variancia aditiva de cada carácter y 1/4 de la covariancia aditiva entre ambos caracteres. En consecuencia, la correlación genética de familias de medios hermanos de X con Y será:

$${}^{MH}r_{Axy} = \frac{1}{4} \sigma_{Ax Ay} / [(\frac{1}{4} \sigma_{Ax}^2)(\frac{1}{4} \sigma_{Ay}^2)]^{0.5},$$

y la correlación genética en la población o entre sus individuos para X con Y será:

$$r_{Axy} = (4) (\frac{1}{4}) \sigma_{Ax Ay} / \{[(4)(\frac{1}{4}) \sigma_{Ax}^2](4)(\frac{1}{4}) \sigma_{Ay}^2\}^{0.5}$$

En el caso de utilizar familias de hermanos completos, la estimación de la correlación genética o aditiva resultará sesgada por la contribución de la dominancia a las variancias y covariancias genotípicas.

11.2.2. Estimación a partir de diseño padre-progenie

El procedimiento requiere la evaluación de los caracteres X e Y en los progenitores y en su descendencia. Suponiendo que se evalúa solo uno de los progenitores (P) y su progenie (H), la correlación genética o aditiva de X con Y será:

$${}^A r_{XY} = \beta_{HXPY} \beta_{HYPX} / \beta_{HXPX} \beta_{HYPY}$$

puesto que:

$$\beta_{\text{HXPY}} = \sigma_{\text{HXPY}} / \sigma_{\text{PY}}^2$$

$$\beta_{\text{HYPX}} = \sigma_{\text{HYPX}} / \sigma_{\text{PX}}^2$$

$$\beta_{\text{HXPX}} = \sigma_{\text{HXPX}} / \sigma_{\text{PX}}^2$$

$$\beta_{\text{HYPY}} = \sigma_{\text{HYPY}} / \sigma_{\text{PY}}^2$$

Simplificando se llega a la expresión:

$${}^A r_{\text{XY}} = \sigma_{\text{HXPY}} \sigma_{\text{HYPX}} / \sigma_{\text{HXPX}} \sigma_{\text{HYPY}}$$

Estas covariancias de padre progenie tienen como esperanza:

$$\sigma_{\text{HXPY}} = 1/2 \sigma_{\text{AXY}}$$

$$\sigma_{\text{HXPX}} = 1/2 \sigma_{\text{AX}}^2$$

$$\sigma_{\text{HYPY}} = 1/2 \sigma_{\text{AY}}^2$$

Finalmente:

$${}^A r_{\text{XY}} = 1/2 \sigma_{\text{AXY}} / [(1/2 \sigma_{\text{AX}}^2)(1/2 \sigma_{\text{AY}}^2)]^{0,5} = \sigma_{\text{AXY}} / \sigma_{\text{AX}} \sigma_{\text{AY}}$$

11.2.3. Estimación a partir de experimentos de selección (o respuesta correlacionada)

Como ya hemos visto en el Capítulo 9, la respuesta a la selección para el carácter X puede predecirse si se conocen la heredabilidad y el diferencial de selección aplicado.

Por lo tanto:

$$R_X = h_x^2 [X_s - X_o]$$

Reemplazando h_x^2 por $\sigma_{\text{AX}} \sigma_{\text{AX}} / \sigma_{\text{PX}} \sigma_{\text{PX}}$

$$R_X = \sigma_{\text{AX}} \sigma_{\text{AX}} [X_s - X_o] / \sigma_{\text{PX}} \sigma_{\text{PX}}$$

O expresando la respuesta empleando la intensidad de selección:

$$R_X = i_x h_x \sigma_{\text{AX}}$$

Si seleccionamos en forma indirecta por el carácter X para mejorar el carácter Y, necesitamos conocer cuál es el cambio en Y por unidad de cambio en X, es decir, el coeficiente de regresión de los valores aditivos de Y en función de las modificaciones por selección en el valor aditivo de X.

Por lo tanto la respuesta correlacionada en Y será:

$$R_{Y.CX} = b_{AXY} R_X,$$

donde

$$b_{AXY} = \sigma_{AXY} / \sigma_{AX}^2$$

Multiplicando el miembro de la derecha de la igualdad por $\sigma_{AY} / \sigma_{AY}$

$$b_{AXY} = \sigma_{AXY} \sigma_{AY} / \sigma_{AX}^2 \sigma_{AY} = r_{AXY} \sigma_{AY} / \sigma_{AX}$$

Finalmente:

$$R_{Y.CX} = b_{AXY} R_X = ({}^A r_{XY} \sigma_{AY} / \sigma_{AX}) (i_X h_X \sigma_{AX}) = i_X h_X {}^A r_{XY} \sigma_{AY}$$

Para expresar la respuesta correlacionada en Y seleccionando por X en función de las heredabilidades, reemplazamos

$$\sigma_{AY} = h_Y \sigma_{PY}$$

$$R_{Y.CX} = i_X {}^A r_{XY} h_X h_Y \sigma_{PY}$$

Por lo tanto la correlación genética o aditiva entre X e Y puede estimarse luego de un experimento de selección midiendo las respuestas directas sobre X y las indirectas sobre Y de acuerdo a la siguiente ecuación:

$${}^A r_{XY} = R_{Y.CX} / i_X h_X h_Y \sigma_{PY}$$

La existencia de correlación entre dos o más características, aunque obedezca a causas genéticas, no puede interpretarse necesariamente como una relación directa de causa-efecto entre los caracteres involucrados. Pero el análisis de las asociaciones estadísticas que ocurren en un grupo de caracteres sí pueden contribuir a modelar las relaciones que podrían existir entre ese grupo de caracteres, y se denomina análisis de coeficientes de sendero (Wright, 1923; Li, 1975). Los coeficientes de sendero permiten descomponer los coeficientes de correlación total entre los caracteres considerados en efectos directos entre pares de caracteres (A y B, a modo de ejemplo) y efectos indirectos causados por la asociación de ambos con otro grupo de caracteres (C, D, etc.). El conjunto de asociaciones posibles entre los caracteres se representa por un diagrama de sendero, cuya estructura, naturaleza y magnitud de asociaciones se aproximan a un modelo causa-efecto determinado empíricamente (Figura 11.4).

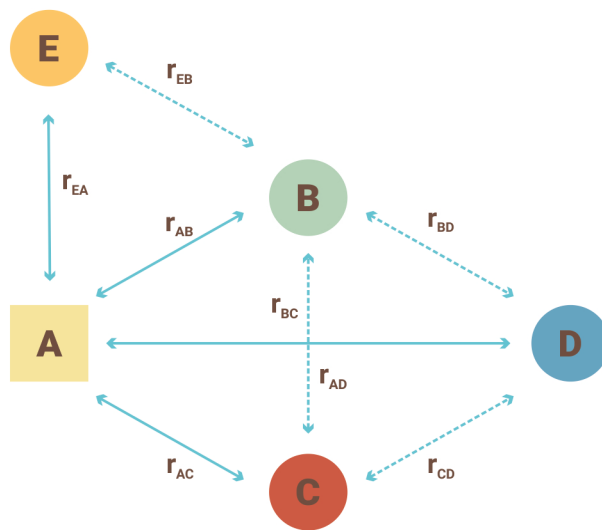


Figura 11.4. Diagrama de sendero hipotético que vincula al carácter A con otros cuatro: B, C, D, E. Las flechas en línea continua indican la correlación total de los caracteres B, C, D y E con A. Las flechas en línea punteada indican las correlaciones totales entre los caracteres B, C, D y E. Obsérvese que E no presenta correlación con C ni D (ver texto).

El diagrama de sendero puede expresar las relaciones de asociación entre un carácter complejo y sus componentes. Un ejemplo puede ser el del rendimiento de grano por planta y sus componentes en maíz, es decir, el número de granos por planta y el peso individual de los granos. A su vez, el número de granos por planta responderá al número de hileras de granos por espiga, al número de granos por hilera y al número de espigas por planta. Como se observa, todo este árbol de asociaciones entre caracteres debe completarse mediante coeficientes de correlación que miden la magnitud de los vínculos entre estos.

Los coeficientes de correlación total entre dos caracteres cualesquiera resultan de la suma de los coeficientes de correlación de los efectos directos entre los dos caracteres y de los efectos indirectos mediados por las asociaciones entre esos dos caracteres con otros caracteres intermediarios. Los efectos directos se denominan coeficientes de sendero y están relacionados con los coeficientes de correlación parcial (ver Anexo), que miden la asociación existente entre dos caracteres manteniendo constantes al resto.

En su versión más sencilla, el análisis de sendero se basa en un análisis de regresión parcial sobre las variables estandarizadas. Como se mencionó en el primer capítulo, el valor de una variable estandarizada surge de restar al valor observado el de su media y expresarlo en desvíos estándar. Así, cada variable, independientemente de su unidad de medida original, seguirá una distribución con media 0 y variancia 1. La covarianza entre variables estandarizadas es numéricamente equivalente a su coeficiente de correlación y a su coeficiente de regresión, puesto que los desvíos estándares y variancias serán todos iguales a 1.

Dado un diagrama de asociación de cuatro caracteres, A, B, C y D, el efecto directo de A sobre B es el coeficiente de regresión parcial entre A y B, manteniendo constantes o excluyendo los efectos de las variables C y D. Es decir, que:

$$r_{AC.BD} = P_{AB}$$

Análogamente:

$$r_{AC.BD} = P_{AC}$$

$$r_{AC.BD} = P_{AD}$$

$$r_{AC.BD} = P_{BC}$$

$$r_{AC.BD} = P_{BD}$$

Si partimos del diagrama de la Figura 11.4, obviando el carácter E para simplificar y se pretende explicar al carácter complejo A en función de los restantes B, C y D, se pueden descomponer las correlaciones observadas entre A y cada uno de ellos formando el sistema de ecuaciones siguiente:

$$r_{BA} = P_{BA} + r_{BC} P_{CA} + r_{BD} P_{DA}$$

$$r_{CA} = r_{CB} P_{BA} + P_{CA} + r_{CD} P_{DA}$$

$$r_{DA} = r_{BD} P_{BA} + r_{DC} P_{CA} + P_{DA}$$

El miembro de la izquierda de la igualdad es un vector de correlaciones observadas, así como los otros coeficientes de correlación que ponderan a los efectos directos también son conocidos. Por lo tanto se trata de un sistema de tres ecuaciones con tres incógnitas, las cuales son los efectos directos de las tres variables, B, C y D, con A (P_{BA} , P_{CA} y P_{DA}).

11.3. Precisión de las estimaciones de correlaciones y relación con el tamaño muestral

Los estadísticos de segundo orden como las variancias de una población están sujetos a grandes errores de muestreo. Tratándose de una población normal, la distribución de variancias muestrales de tamaño n en relación con la variancia paramétrica poblacional sigue una distribución Chi cuadrado (χ^2) con n-1 grados de libertad:

$$(n-1)S^2/\sigma^2 \sim \chi^2_{\alpha,(n-1)}$$

Esta relación puede utilizarse para construir intervalos de confianza para la relación entre la variancia muestral y la variancia paramétrica y para las correlaciones (Baker, 1986) (Figuras 11.5 y 11.6).

En el caso de las correlaciones, el grado de precisión guarda relación no solo con el de las estimaciones de las variancias, sino también con el de las covariancias entre caracteres. Cuanto mayor sea el número de variancias y covariancias que se necesiten estimar simultáneamente, menores serán las precisiones alcanzables. La variancia del estimador de covariancia entre dos caracteres, X e Y es:

$$1/n [(1 + r^2)\sigma_X^2 \sigma_Y^2]$$

Es decir, que los errores de muestreo de estimaciones de covariancias poblacionales tenderán a ser menores cuanto mayor sea la correlación entre los caracteres (Tallis, 1959).

Nótese que la obtención de estimadores de correlación que no se desvíen más de 0,1 a 0,2 unidades respecto de la correlación paramétrica requiere tamaños muestrales (n) entre 300 y 400 (Figura 11.6). Los requerimientos de tamaño muestral tienden a ser mayores cuanto más reducido sea el coeficiente de correlación paramétrico.

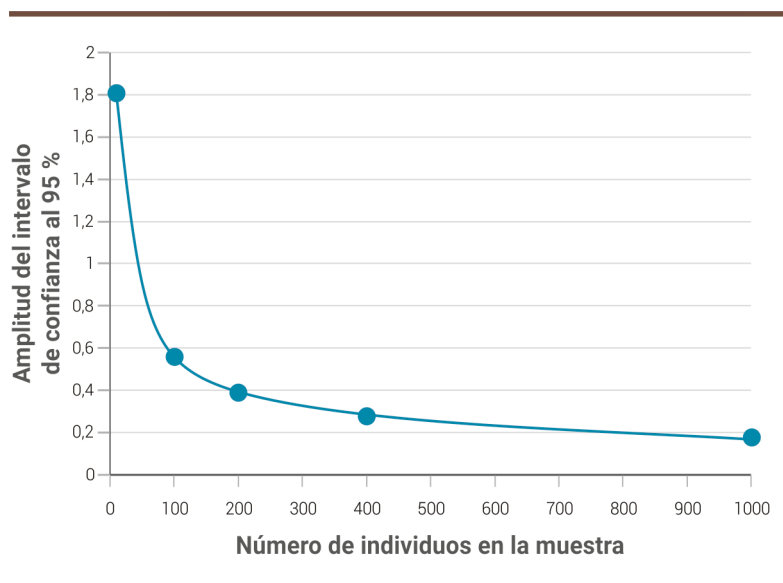


Figura 11.5. Amplitud del intervalo de confianza para la estimación de la variancia paramétrica poblacional al 95 %, a partir de muestras de diferente tamaño. Basado en Baker, 1986.

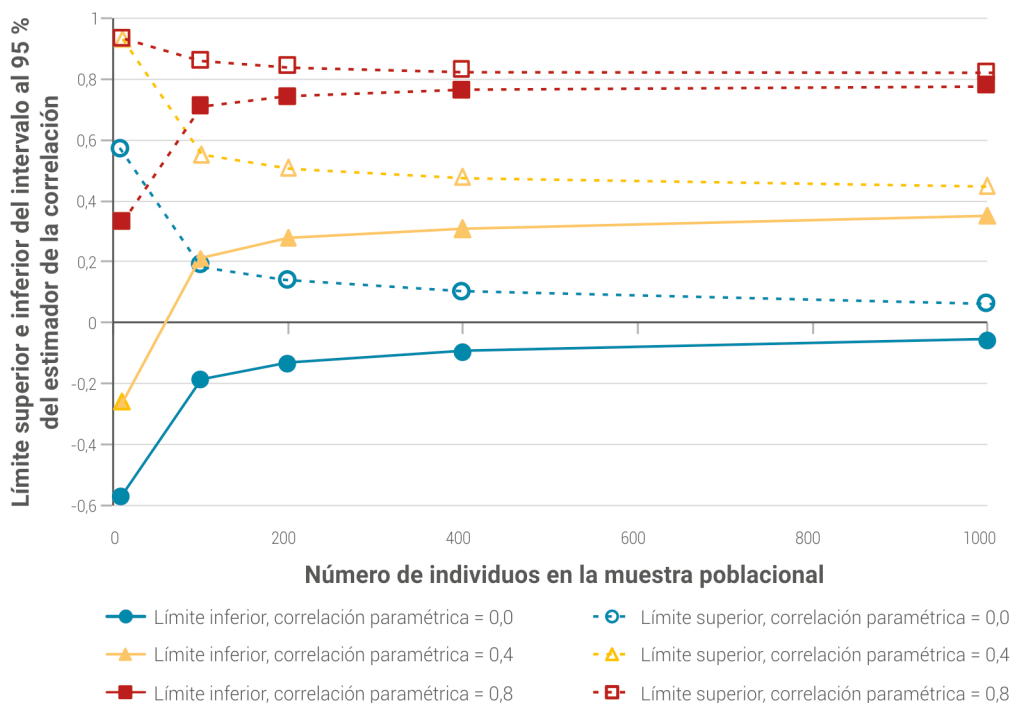


Figura 11.6.

Límites superiores e inferiores del intervalo de confianza (95 %) para las estimaciones muestrales de la correlación entre caracteres, asumiendo que los verdaderos coeficientes de correlación poblacional sean 0,0; 0,4 y 0,8 para diferentes tamaños de muestra. Basado en Baker, 1986

11.4. Correlación genética entre resultados de evaluaciones de genotipos en etapas tempranas y tardías

Bernardo (1992) demostró que las ventajas de realizar pruebas tempranas de habilidad combinatoria de líneas endocriadas disminuyen en la medida que la heredabilidad de las medias de cruzamientos de prueba (*test crosses*) en esas etapas sea baja. El objetivo de las evaluaciones tempranas de habilidad combinatoria es realizar un descarte inicial de genotipos que deje disponible mayor cantidad de recursos (en términos de parcelas de evaluación, repeticiones y ambientes) para evaluar con mayor heredabilidad al resto de los genotipos presuntamente más promisorios. Este descarte temprano, sin embargo, conlleva el riesgo de eliminar genotipos cuyo mérito genético, siendo superior, no se manifestó inicialmente a causa de factores ambientales o deficiencias experimentales. Para reducir ese riesgo el fitomejorador puede decidir trabajar con un mayor nivel de seguridad eliminando una menor proporción de líneas, pero de ser así es más probable

que se asignen parte de los recursos limitados a reiterar la evaluación de genotipos que realmente no poseerían el mérito genético buscado, y se los retacee para evaluar a los genotipos verdaderamente promisorios.

Por ello resulta conveniente disponer de una guía orientativa que ayude a tomar la decisión sobre qué porcentaje o proporción de líneas convendría retener en las etapas iniciales. Para ello es importante conocer las correlaciones genéticas y fenotípicas entre cruzamientos de prueba de líneas con diferentes niveles de endocría y calcular la intensidad de selección para aplicar en las diferentes etapas.

Consideremos que se parte de una población de cría en equilibrio de Hardy-Weinberg y que se desarrollan líneas endocriadas de manera aleatoria (sin selección), que las acciones génicas involucradas son las de aditividad y dominancia, y ausencia de epistasia. En cada generación de endocría, las líneas son cruzadas por un probador común o *tester* y los cruzamientos de prueba se evalúan en experimentos. En tal caso nos interesa determinar la probabilidad de retener (seleccionar) las líneas genéticamente superiores en las primeras etapas de evaluación. Esa probabilidad depende de la correlación genética entre el comportamiento de los cruzamientos de prueba de, por ejemplo, una filial con endocría temprana y otra correspondiente al final del proceso de endocría, a las que llamaremos generaciones t y t' .

Bernardo (1992) demostró que esa correlación es:

$$r_{tt'} = [(1 + F_t)/(1 + F_{t'})]^{0.5}$$

Donde F es el coeficiente de endocría, cuyo valor será

$F_t = 1 - (1/2)^{t-1}$ cuando la semilla de varias plantas de la filial t se cosecha o maneja posteriormente en *bulk*.

En caso de que se tomara la semilla de una planta autofecundada en la filial t ,

$$F_t = 1 - (1/2)^t$$

Como puede observarse (Tabla 11.3) existe un elevado nivel de correlación genética entre el comportamiento de los cruzamientos de prueba realizados en filiales tempranas y tardías, por lo cual es de esperar que las pruebas en filiales tempranas tengan una alta eficiencia para seleccionar las mejores líneas en etapas tempranas. Estas correlaciones aumentan cuanto mayor sea la heredabilidad de los cruzamientos de prueba en la población original y cuanto mayor sea el nivel de endocría de las filiales que se consideren.

Filiales	Niveles de endocría de las líneas					
	S ₂	S ₃	S ₄	S ₅	S ₆	S _n
S ₁	0,82	0,76	0,73	0,72	0,71	0,71
S ₂		0,93	0,89	0,88	0,87	0,87
S ₃			0,97	0,95	0,94	0,94
S ₄				0,98	0,98	0,97
S ₅					0,99	0,98
S ₆						0,99

Tabla 11.3. Correlaciones genéticas entre el comportamiento de cruzamientos de prueba de líneas en diferentes niveles de endocría. Adaptado de Bernardo, 1992.

La correlación que nos interesa conocer es la que exista entre los valores fenotípicos de los cruzamientos de prueba de líneas en la generación t, con los verdaderos méritos o valores genéticos de esas mismas líneas en cruzamientos de prueba una vez que alcanzaron el nivel de homocigosis (t_∞). Es decir:

$$r_{P_t, G_{t\infty}} = r_{G_t, t\infty} h_t = [(1 + F_t)/2]^{0,5} h_t$$

donde h_t es la raíz cuadrada de la heredabilidad de los cruzamientos de prueba de las líneas en la generación de endocría t.

Las correlaciones entre los valores fenotípicos de los cruzamientos de prueba de líneas con diferentes niveles de endocría con el valor genotípico verdadero de los cruzamientos de prueba de las mismas líneas en estado homocigota se presentan en la Figura 11.7.

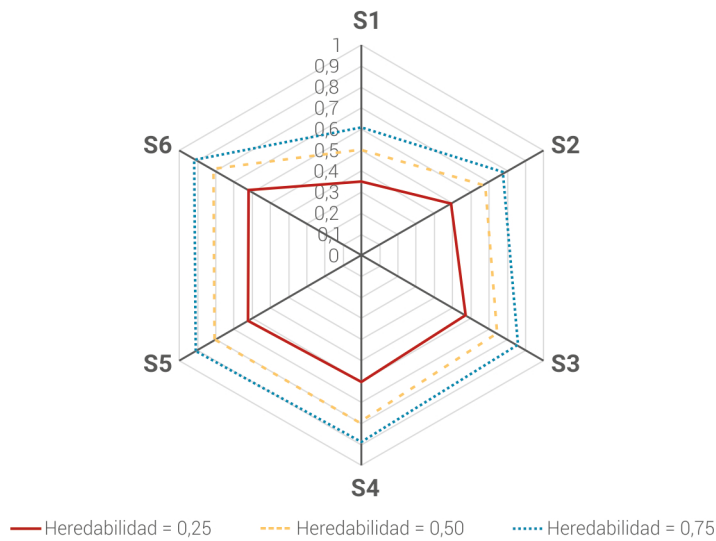
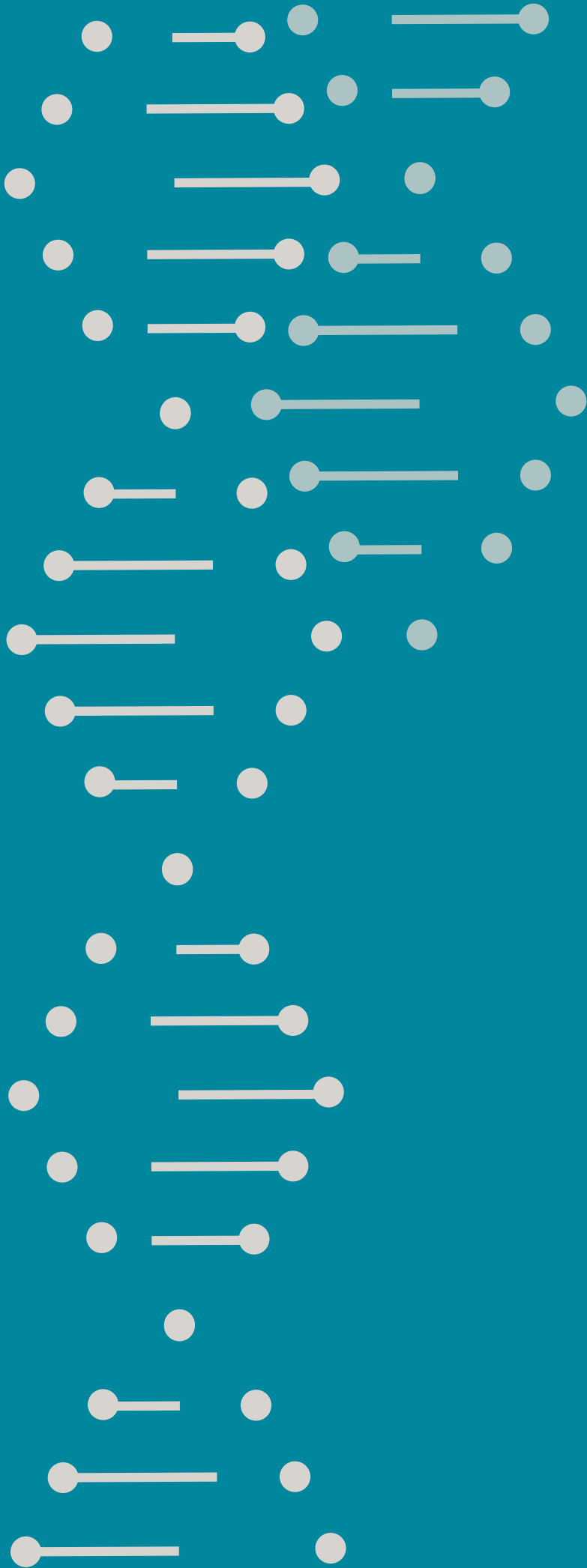


Figura 11.7. Correlaciones entre los valores fenotípicos de los cruzamientos de prueba en diferentes niveles de endocria de las líneas evaluadas (S1 a S6) y el valor genotípico de los cruzamientos de prueba de las líneas al estado homocigota para tres niveles de heredabilidad. Basado en Bernardo, 1992.

Con esa información, para niveles arbitrarios de heredabilidad y de probabilidad de retener a aquellas que luego serán genéticamente superiores una vez estabilizadas en homocigosis (con base en los cruzamientos de prueba de líneas en la generación t de endocria) Bernardo (1992) presentó una serie de tablas para determinar el porcentaje de líneas que convendría retener. Estas tablas presentan la probabilidad condicional de que el valor genético de un cruzamiento de prueba de una línea homocigota (etapa t_{∞}) esté en la fracción superior de la distribución (α_{∞}), dado que el valor fenotípico del cruzamiento de prueba de la misma línea haya estado en la etapa t también en la fracción superior de la distribución (α_t).





Selección de múltiples caracteres

Existen tres maneras principales de practicar selección cuando se tienen en cuenta más de un carácter cuantitativo:

- i. En tándem.
- ii. En niveles de corte independientes.
- iii. Mediante índices de selección.

12.1. Selección en tándem

La selección en tándem (Hazel y Lush, 1942) para varios caracteres no es una selección simultánea sino secuencial. Consiste en realizar uno o más ciclos de selección para el carácter X y una vez alcanzada la respues-

ta deseada para X, se comienza a seleccionar por Y y así sucesivamente si existiera un tercer carácter de interés. Para cada etapa selectiva, por lo tanto, se utiliza un único punto de truncamiento. En la medida que los caracteres considerados estén genéticamente correlacionados, en cada etapa pueden esperarse respuestas indirectas en el mismo sentido de la correlación (ver Figura 11.1).

Si la selección para un carácter X resultó eficiente y este carácter está genéticamente correlacionado con otro carácter Y, provocará cambios (disminución) en la variancia genética remanente del carácter Y, $\Delta(^A\sigma_Y^2)$. Esa reducción dependerá de la intensidad de selección aplicada en el carácter X y de la correlación genética entre X e Y.

$$\Delta(^A\sigma_Y^2) = (^A\sigma_Y^2)[1 - ^Ar_{XY}^2(z/p)[(z/p) - t]],$$

donde (Figura 12.1),

$^Ar_{XY}$ es la correlación genética entre los caracteres X e Y,

z es la ordenada al origen en el punto de truncamiento T,

t es el desvío entre el punto de truncamiento y la media poblacional expresado en desvíos estándar,

p es la proporción de individuos o familias seleccionados para X.

Existen tablas (Tabla 12.1) donde se tabulan los cambios esperados en Y con relación a los que ocurren en X basadas en las propiedades de la distribución normal (Freeman, 1988).

Porcentaje seleccionado con base en X	Intensidad de selección	Proporción de variancia aditiva remanente en el carácter Y para distintos coeficientes de correlación genética entre X e Y			
		1,0	0,75	0,50	0,25
(p)	(i)				
1	2.665	0,097	0,492	0,774	0,944
5	2.063	0,138	0,515	0,785	0,946
10	1.755	0,169	0,533	0,792	0,948
20	1.400	0,219	0,561	0,805	0,951
30	1.159	0,265	0,586	0,816	0,954
50	0,798	0,363	0,642	0,841	0,960

Tabla 12.1

Afectación relativa de la variancia aditiva de un carácter Y cuando se practica selección sobre el carácter X, considerando diferentes correlaciones genéticas entre X e Y. Adaptado de Freeman, 1988.

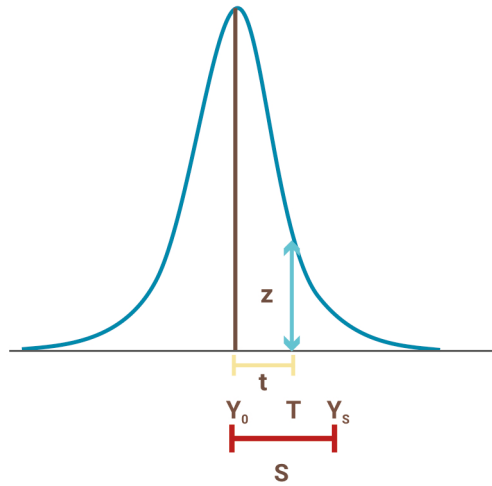


Figura 12.1. Selección para un carácter Y a partir de un punto de truncamiento T. Y_0 e Y_s indican la media poblacional y la media de los individuos seleccionados. El segmento z es la ordenada a la curva de distribución de frecuencia en el punto de truncamiento y t es la distancia medida en unidades estándar entre el punto de truncamiento T y la media poblacional Y_0 .

12.2. Selección por niveles de cortes independientes

En este tipo de selección se utilizan tantos puntos de truncamiento T_j como caracteres se van a seleccionar en el mismo ciclo. A los puntos de truncamiento se los denomina niveles de corte y son independientes por cuanto son fijados *a priori* por el fitomejorador basado en su conocimiento y los objetivos del programa (Figura 12.2).

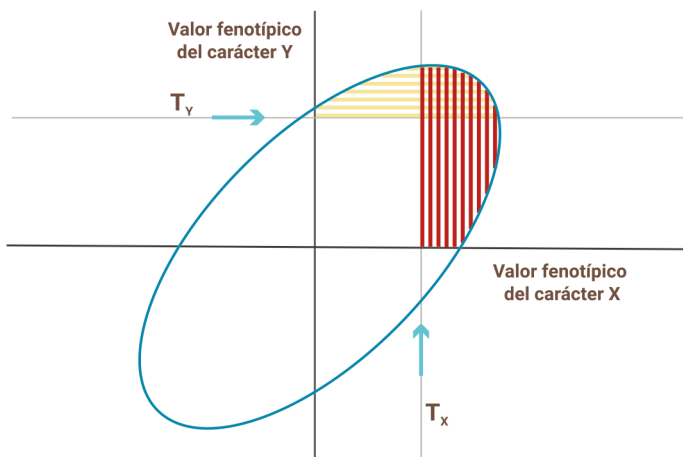


Figura 12.2. Selección por niveles de corte independientes para una población considerando dos caracteres positivamente correlacionados. T_y y T_x son los puntos de truncamiento; α_x y α_y son las fracciones seleccionadas por cada punto de truncamiento (áreas de simple rayado en color rojo y amarillo, respectivamente). La fracción finalmente seleccionada (α) es la fracción del gráfico que superpone a las dos anteriores y se presenta con doble rayado en color.

Es decir, que al definir un valor arbitrario de T_x y T_y la fracción α ya queda determinada ($\alpha = \alpha_x \alpha_y$). El tamaño o fracción seleccionada de la población de-

pende no solo de la ubicación o de la posición del punto de truncamiento en la distribución fenotípica de cada carácter, sino también de la correlación fenotípica (en magnitud y signo) existente entre los dos caracteres (Figura 12.3).

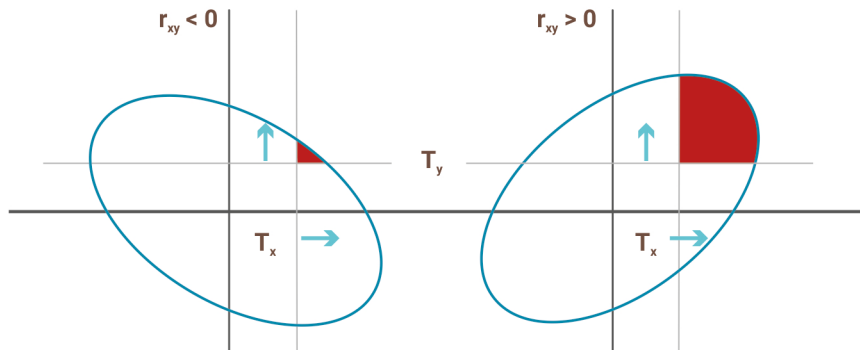


Figura 12.3. Selección por niveles de corte independientes para dos poblaciones hipotéticas o pares de caracteres positiva y negativamente correlacionados.

Cuando la selección es por dos niveles de corte independiente (caracteres X e Y), la respuesta a la selección para el j-ésimo carácter incluido en la selección será

$$R_j = (1/\alpha) [r_{p_xG_j} z_1 I_2 + r_{p_yG_j} z_2 I_1] \text{ (Wricke y Weber, 1986)}$$

$r_{p_xG_j}$ es la correlación entre el valor fenotípico del carácter X y el valor genotípico del carácter j-ésimo; $r_{p_yG_j}$ es la correlación entre el valor fenotípico del carácter Y y el genotípico para el carácter j-ésimo; z_1 y z_2 son las ordenadas al origen en los puntos de truncamiento para X y para Y expresados en desvíos estándares; e I_1 e I_2 son funciones de la correlación fenotípica entre los dos caracteres ($\rho_{p_xp_y}$) (Wricke y Weber, 1986).

Cuando los caracteres no están correlacionados:

$$R_j = i_{\alpha_j} r_{p_jG_j}$$

donde

i_{α_j} es la intensidad de selección correspondiente a la proporción α de individuos seleccionados para el carácter j-ésimo y $r_{p_jG_j}$ es la correlación entre los valores fenotípicos y genotípicos para el j-ésimo carácter.

12.3. Índices de selección

Los índices de selección son funciones lineales de los valores fenotípicos para varios caracteres (o para un carácter, pero provenientes de distintas fuentes de información, por ejemplo de dos tipos de parientes), donde cada uno de esos valores recibe una ponderación denominada peso del índice. A los índices de selección se los denomina también índices fenotípicos porque generalmente se construyen con base en medidas fenotípicas y se busca que constituyan la mejor predicción lineal del valor de mejora de las unidades de selección considerando simultáneamente los caracteres incluidos en el índice (Figura 12.4).

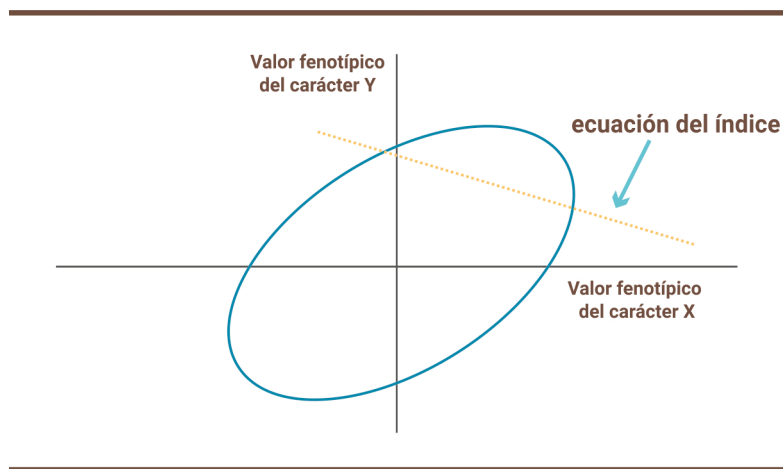


Figura 12.4. Índice de selección fenotípica construido para dos caracteres positivamente correlacionados.

El concepto de los índices de selección se debe a los aportes de Smith (1936) y de Fisher (1936) y pueden aplicarse con dos objetivos generales de mejoramiento:

- a. Mejoramiento de un único carácter (como caso especial):
 - combinando información de diferentes etapas de evaluación;
 - combinando información de parientes en una misma evaluación.
- b. Mejoramiento simultáneo de dos o más caracteres.

Si bien los índices de selección fueron inicialmente más adoptados para el mejoramiento animal, han alcanzado una creciente atención en los programas de mejoramiento vegetal. De hecho, y según las circunstancias, para la mejora simultánea de más de un carácter, los índices de selección resultarían igualmente o más eficientes que la selección por niveles de cortes independientes o la selección en tándem. Los índices de selección han probado ser muy útiles en programas de mejoramiento a largo plazo basados en selección recurrente. En el caso de programas de más corto plazo (a los que se les dedican pocos ciclos de selección) su ventaja se reduce. No

obstante en estas últimas situaciones pueden ser muy eficientes porque proveen un criterio objetivo para la elección de los mejores progenitores que servirán como material de partida para generar poblaciones de cría. En especies autógamias, los índices de selección se aplican también en la selección en dos o más etapas (*multistage selection*).

Las medidas fenotípicas tomadas sobre las unidades de selección para varios caracteres pueden incorporarse en el índice de diversas maneras, como por ejemplo mediante una combinación lineal del siguiente tipo:

$$I = b_1 p_1 + b_2 p_2 + \dots + b_n p_n$$

En esta ecuación del índice (I), los coeficientes b son los pesos del índice que ponderan a los valores fenotípicos p, para los n caracteres incluidos en el índice.

De manera general, lo expresado más arriba equivale a una sumatoria de n términos:

$$I = \sum_{i=1}^n b_i$$

Los valores fenotípicos pueden estar expresados en las unidades originales de cada variable, o expresarse en forma estandarizada.

Con la aplicación de un índice de selección (I) se desea obtener la mayor respuesta posible en el genotipo agregado. El genotipo agregado o índice genotípico H es una medida del valor genotípico combinado a través de varios caracteres. Los pesos del índice I se estiman de modo tal de maximizar la respuesta en H. La maximización de la respuesta se logrará en la medida en que el índice fenotípico I resulte ser la mejor predicción lineal del valor de mejora agregado o compuesto (para varios caracteres) H de un individuo.

El genotipo agregado es una función lineal de los valores genotípicos (no observables), ponderados por sus respectivos valores económicos relativos, los cuales se conocen o fijan arbitrariamente (Baker, 1986). Esta ponderación está determinada por el interés del fitomejorador en función de los objetivos del programa de mejoramiento y las características de la población en la que practicará la selección. El genotipo agregado o índice genotípico puede expresarse mediante una combinación lineal análoga al índice fenotípico de la siguiente forma:

$$H = a_1 g_1 + a_2 g_2 + \dots + a_m g_m$$

O bien:

$$H = \sum_{j=1}^m a_j g_j$$



donde

g_j denota los valores de mejora (o valores reproductivos) de cada individuo,

a_j corresponde al peso económico, o peso relativo de cada carácter que se pretende mejorar.

m es el número de caracteres para mejorar.

El signo del peso económico del carácter indica si el carácter contribuye en forma favorable (si es positivo) o en forma desfavorable (si es negativo) al mérito agregado.

Baker (1986) sostiene que los pesos económicos deben establecerse respetando la proporcionalidad del efecto de los cambios en uno de los caracteres sobre los restantes. Dados tres caracteres, A, B y C, si conociéramos que un incremento de 5 unidades en el carácter A equivalieran a una incidencia de 10 unidades en B y 20 en C:

$$5 A \equiv 10 B \equiv 20 C$$

Los pesos económicos deberían ser:

$$a_A = 5/5 = 1$$

$$a_B = 5/10 = 0,50$$

$$a_C = 5/20 = 0,25$$

de modo tal que:

$$a' = (1 \quad 0,50 \quad 0,25)$$

En relación con el mismo tema, Cruz y Regazzi (1994) proponen utilizar la relación entre los coeficientes de variación genotípicos.

Es importante tener en cuenta que hay n caracteres para medir (por su fenotipo) y m caracteres para mejorar. Téngase en cuenta que este último número puede ser igual o menor que n . Al aplicar la teoría de índices de selección, se asume que se dispone de la misma cantidad de información en todas las unidades de selección.

Para generar un índice de selección que permita lograr la respuesta esperada hace falta disponer de estimaciones de parámetros poblacionales que sean lo más precisas posibles. La precisión en los estimadores está en relación inversa con el tamaño de la muestra de genotipos tomada de



la población, pero también disminuye cuanto mayor es el número de caracteres que se pretende incorporar en el índice y si algunos de ellos están altamente correlacionados entre sí.

Teniendo en cuenta los factores que afectan la precisión de los estimadores de parámetros poblacionales, las pautas generales que deberían guiar la decisión sobre cuáles y cuántos caracteres deben incorporarse en el índice de selección son las siguientes:

- excluir del índice aquellos caracteres que tengan una baja correlación parcial con el genotipo agregado;
- excluir los caracteres que provean información que resulte redundante con la provista por otros caracteres ya incorporados en el índice y que por lo tanto no contribuirán de manera significativa al mejoramiento del genotipo agregado;
- ciertos caracteres surgen de la manipulación matemática de otro. Por tal motivo, en estos casos debería elegirse solo uno de los caracteres que muestren elevada correlación.

Independientemente de los criterios señalados es casi inevitable cierta subjetividad en la decisión sobre cuáles caracteres se utilizarán en el índice.

12.3.1. Ecuaciones del índice de selección

Aquellos pesos del índice (b_j) que maximizan la respuesta en H se denominan, por tal razón, pesos óptimos. La estimación de los pesos del índice se hace a partir de la resolución de esas incógnitas a partir de sistemas de ecuaciones del tipo siguiente [12.1]:

$$b_1^P \sigma_1^2 + b_2^P \sigma_{12} + b_3^P \sigma_{13} + \dots + b_n^P \sigma_{1n} = a_1^G \sigma_1^2 + a_2^G \sigma_{12} + a_3^G \sigma_{13} + \dots + a_m^G \sigma_{1m}$$

$$b_1^P \sigma_{21} + b_2^P \sigma_2^2 + b_3^P \sigma_{23} + \dots + b_n^P \sigma_{2n} = a_1^G \sigma_{21} + a_2^G \sigma_2^2 + a_3^G \sigma_{23} + \dots + a_m^G \sigma_{2m}$$

$$b_1^P \sigma_{31} + b_2^P \sigma_{32} + b_3^P \sigma_3^2 + \dots + b_n^P \sigma_{3n} = a_1^G \sigma_{31} + a_2^G \sigma_{32} + a_3^G \sigma_3^2 + \dots + a_m^G \sigma_{3m}$$

.....

$$b_1^P \sigma_{n1} + b_2^P \sigma_{n2} + b_3^P \sigma_{n3} + \dots + b_n^P \sigma_n^2 = a_1^G \sigma_{n1} + a_2^G \sigma_{n2} + a_3^G \sigma_{n3} + \dots + a_m^G \sigma_n^2$$

donde

σ_i^2 y σ_{ij} es la variancia fenotípica del carácter i-ésimo y la covariancia fenotípica entre los caracteres i y j, respectivamente;

${}^G\sigma_i^2$ y ${}^G\sigma_{ij}$ es la variancia genotípica del carácter i-ésimo y la covariancia genotípica entre los caracteres i y j, respectivamente (Baker, 1986).

Siguiendo la terminología utilizada para selección para un único carácter, el índice de selección I es el criterio de selección y el objetivo de selección es mejorar el genotipo o mérito agregado H.

12.3.2. Propiedades del índice y del genotipo agregado

Se asume que I y H tienen distribución normal multivariada y que por lo tanto puede encontrarse una regresión lineal de H en función de I. La regresión, sin embargo, puede existir aun sin la condición de normalidad (Baker, 1986).

Cuando se cumple el supuesto de que las distribuciones de los valores genotípicos y fenotípicos son normales, se obtendría la mayor respuesta posible al índice, además de que existirá una relación simple entre la intensidad de selección y el diferencial de selección estandarizado, útil para predecir más fácilmente la respuesta a la selección.

Si bien la violación del supuesto de normalidad no invalida la aplicación práctica de los índices de selección, no es factible estimar correctamente las modificaciones que, por efecto de la aplicación del índice, ocurran en las variancias y covariancias genotípicas. Cuando existen evidencias fuertes de desvíos de la normalidad o linealidad de los efectos, debe tomarse la decisión de transformar los datos para alcanzar esa condición, o como sugiere Baker (1986), emplear directamente índices no lineales.

Las esperanzas del genotipo agregado y las del índice fenotípico serán, respectivamente:

$$\varepsilon(H) = \sum a_i \varepsilon(g_i) = 0$$

$$\varepsilon(I) = \sum b_j \varepsilon(p_j) = 0$$

Las variancias del genotipo agregado y el índice son:

$$V(H) = \sigma_H^2 = \sum \sum a_i a_j {}^G\sigma_{ij}$$

$$V(I) = \sigma_I^2 = \sum \sum b_i b_j {}^P\sigma_{ij}$$

La covariancia entre el índice fenotípico y el genotipo agregado es:

$$\text{Cov}(HI) = \sigma_{HI} = \sum \sum a_i b_j {}^G\sigma_{ij}$$

El término ${}^G\sigma_{ij}$ hace referencia a la covariancia de los valores genotípicos entre los diferentes caracteres incluidos en el genotipo agregado o a las variancias genotípicas (si $i=j$); mientras que el término ${}^P\sigma_{ij}$ corresponde a las covariancias fenotípicas de los caracteres incluidos en el índice fenotípico o las variancias fenotípicas (si $i=j$).

A manera de ejemplo, las variancias del índice fenotípico y del genotipo agregado construido con dos caracteres, denominados 1 y 2, tales que

$$I = b_1p_1 + b_2p_2$$

$$H = a_1g_1 + a_2g_2$$

serán

$$\sigma_I^2 = V(b_1p_1 + b_2p_2) = (b_1p_1 + b_2p_2)^2 = b_1^2p_1^2 + b_2^2p_2^2 + 2b_1b_2p_1p_2 = b_1^2{}^P\sigma_1^2 + b_2^2{}^P\sigma_2^2 + 2b_1b_2{}^P\sigma_{12} \text{ (ver variancia de una función lineal en Anexo)}$$

$$\sigma_H^2 = V(a_1g_1 + a_2g_2) = (a_1g_1 + a_2g_2)^2 = a_1^2g_1^2 + a_2^2g_2^2 + 2a_1a_2g_1g_2 = a_1^2{}^G\sigma_1^2 + a_2^2{}^G\sigma_2^2 + 2a_1a_2{}^G\sigma_{12}$$

$$\sigma_{HI} = \text{COV}(b_1p_1 + b_2p_2, a_1g_1 + a_2g_2) = a_1b_1{}^G\sigma_1^2 + a_2b_1{}^G\sigma_{12} + a_1b_2{}^G\sigma_{12} + a_2b_2{}^G\sigma_2^2$$

A su vez considerando el sistema de ecuaciones [12.1]:

$$\sum b_i{}^P\sigma_{ij} = \sum a_i{}^G\sigma_{ij}$$

aplicado al mismo índice tendremos el siguiente sistema:

$$b_1{}^P\sigma_1^2 + b_2{}^P\sigma_{12} = a_1{}^G\sigma_1^2 + a_2{}^G\sigma_{12}$$

$$b_1{}^P\sigma_{21} + b_2{}^P\sigma_2^2 = a_1{}^G\sigma_{21} + a_2{}^G\sigma_2^2$$

Si tenemos en cuenta que

$$\sigma_{HI} = \sum \sum a_i b_j{}^G\sigma_{ij} = \sum b_j \sum a_i{}^G\sigma_{ij}$$

$$\begin{aligned} \text{entonces } \sigma_{HI} &= b_1(a_1{}^G\sigma_1^2 + a_2{}^G\sigma_{12}) + b_2(a_1{}^G\sigma_{12} + a_2{}^G\sigma_2^2) \\ &= b_1(b_1{}^P\sigma_1^2 + b_2{}^P\sigma_{12}) + b_2(b_1{}^P\sigma_{21} + b_2{}^P\sigma_2^2) \\ &= b_1^2{}^P\sigma_1^2 + b_2^2{}^P\sigma_2^2 + 2b_1b_2{}^P\sigma_{21} = \sigma_I^2 \end{aligned}$$

La igualdad $\sigma_{HI} = \sigma_I^2$ se cumple a condición de que la cantidad de información disponible para los caracteres incorporados en el índice sea la misma.

La correlación entre el índice y el genotipo agregado (r_{HI}) es:

$$r_{HI} = \sigma_{HI} / \sigma_I \sigma_H = \sum \sum a_i b_j^G \sigma_{ij} / [(\sum \sum a_i a_j^G \sigma_{ij}) (\sum \sum b_i b_j^P \sigma_{ij})]^{0.5}$$

y la covarianza entre el valor genotípico del carácter i -ésimo y el índice será:

$$COV(G_i I) = \sigma_{(G_i I)} = \sum_i b_j^G \sigma_{ij}$$

12.3.3. Respuesta a la selección en el genotipo agregado

Si las variables medidas siguen una distribución normal, también lo hará el índice. Por lo tanto, la respuesta a la selección mediante índices puede predecirse de la misma manera que la selección por truncamiento para un único carácter (Figuras 12.5 y 12.6).

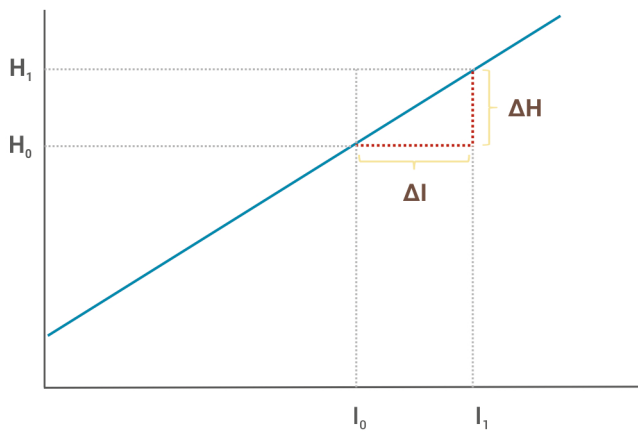


Figura 12.5. Respuesta a la selección (ΔH) por un índice fenotípico. I_0 e I_s son los valores promedio del índice fenotípico para la población original y de las unidades seleccionadas, respectivamente.

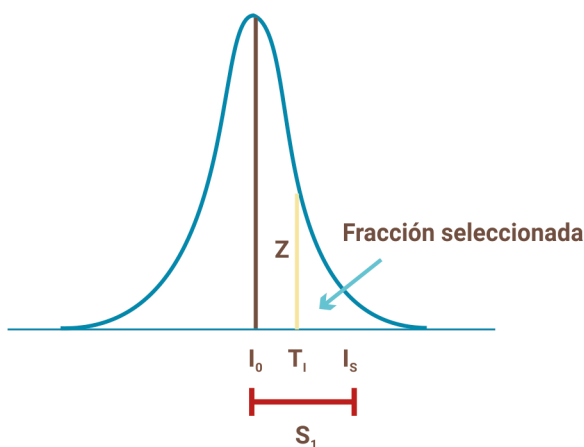


Figura 12.6. Selección por truncamiento aplicada a una población cuyo índice fenotípico sigue la distribución normal. Z es la ordenada en el punto de truncamiento (T_1) a partir del cual se selecciona la proporción p de genotipos.

La respuesta a la selección o el cambio en el mérito agregado será entonces

$$R = \Delta H = H_s - H_o = b_{HI} (I_s - I_o),$$

donde

$(I_s - I_o)$ es el diferencial de selección o diferencia entre el valor promedio del índice en la población inicial y las unidades seleccionadas,

b_{HI} es la pendiente o cambio en el mérito agregado por unidad de cambio en el valor observado para el índice

$$b_{HI} = \sigma_{HI} / \sigma_I^2$$

Por lo tanto:

$$\Delta H = (\sigma_{HI} / \sigma_I^2) (I_s - I_o) \quad [12.2]$$

Si se dividen ambos miembros de la igualdad por el desvío estándar del genotipo agregado:

$$\Delta H / \sigma_H = [(\sigma_{HI} / \sigma_I^2) (I_s - I_o)] / \sigma_H = (\sigma_{HI} / \sigma_I \sigma_H) (I_s - I_o) / \sigma_I \quad [12.3]$$

La intensidad de selección por truncamiento es

$$i = (I_s - I_o) / \sigma_I \quad [12.4]$$

Reemplazando [12.4] en [12.3] y despejando ΔH obtendremos que:

$$\Delta H = i r_{HI} \sigma_H \quad [12.5]$$

Si en cambio se dividen ambos miembros de la igualdad [12.2] por el desvío estándar del índice fenotípico:

$$\Delta H / \sigma_I = [(\sigma_{HI} / \sigma_I^2) (I_s - I_o)] / \sigma_I = b_{HI} (I_s - I_o) / \sigma_I \quad [12.6]$$

Entonces

$$\Delta H = i b_{HI} \sigma_I \quad [12.7]$$

Observemos la relación entre r_{HI} y b_{HI} :

$$r_{HI} = \sigma_{HI} / \sigma_I \sigma_H$$

$$b_{HI} = \sigma_{HI} / \sigma_I \sigma_I$$

Por lo tanto:

$$r_{HI} = b_{HI} \sigma_I / \sigma_H \quad [12.8]$$

Esta correlación entre el mérito agregado y el fenotipo para el índice se denomina precisión de selección (en inglés, *accuracy of selection*). Depende de cuánta información está disponible y de cómo se la utiliza. La precisión de selección aumentará cuantas más observaciones haya de los genotipos (a través de repeticiones y de ambientes) y cuanto menos importante resulte la contribución de la correlación ambiental a la correlación fenotípica. Recuérdese que la contribución de la correlación genética a la correlación fenotípica entre dos caracteres será mayor cuanto mayor sea el producto de las raíces cuadradas de las respectivas heredabilidades.

Tomando en cuenta las expresiones [12.5] y [12.8], cuanto mayor sea la precisión de selección mayor será la respuesta. Como σ_H es un parámetro poblacional, para mejorar la respuesta solo podemos accionar sobre la magnitud del producto $b_{HI} \sigma_I$ maximizándolo.

Otra expresión equivalente a [12.5] y [12.7] de la respuesta esperada a la selección utilizando un índice fenotípico es la siguiente:

$$\Delta H = i r_{HI} \sigma_H = i b_{HI} \sigma_I = i \sigma_{HI} / \sigma_I = i \sigma_I, \quad [12.9]$$

donde

r_{HI} es la precisión de selección, o correlación entre el índice y el genotipo agregado.

σ_{HI} es la covariancia entre el índice fenotípico y el genotipo agregado.

σ_I y σ_H son los desvíos estándares del índice y del genotipo agregado, respectivamente.

Los pesos óptimos del índice (b_i) serán aquellos que maximicen el valor absoluto de la covariancia σ_{HI} y estadísticamente son los coeficientes de regresión parcial de los valores fenotípicos (p_i) en función del genotipo agregado H. Además de poder estimar la respuesta en el genotipo agregado, es también posible hacerlo para cada carácter individual incluido en H.

Esta será:

$$\Delta G_i = R_i = i b_{GiI} \sigma_I = i r_{GiI} ({}^G\sigma_i^2)^{0.5}, \quad [10.12]$$

donde:

b_{GiI} es el coeficiente de regresión entre el valor genotípico para el carácter i-ésimo y el valor fenotípico del índice.

r_{GiI} es el coeficiente de correlación entre el valor genotípico para el carácter i-ésimo y el valor fenotípico del índice.

${}^G\sigma_i^2$ es la variancia genotípica para el carácter i-ésimo.

De manera similar, la respuesta medida en términos del valor fenotípico del índice será

$$\Delta P_i = i b_{piI} \sigma_i = i r_{piI} ({}^P\sigma_i^2)^{0.5}, \quad [12.11]$$

donde:

b_{piI} es el coeficiente de regresión entre el valor fenotípico para el carácter i -ésimo y el valor fenotípico del índice.

r_{piI} es el coeficiente de correlación entre el valor fenotípico para el carácter i -ésimo y el valor fenotípico del índice.

${}^P\sigma_i^2$ es la variancia fenotípica para el carácter i -ésimo.

Los pesos óptimos dependen de:

- las variancias fenotípicas de los caracteres incluidos en el índice;
- las covariancias fenotípicas entre los caracteres incluidos en el índice;
- la covariancia entre los valores fenotípicos de los caracteres incluidos en el índice y el genotipo agregado.

12.3.4. Expresión de los índices en términos matriciales

El sistema de ecuaciones para estimar los pesos del índice presentado más arriba [12.1] puede expresarse mediante álgebra matricial de la siguiente manera:

$$P\mathbf{b} = G\mathbf{a} \quad [12.12]$$

donde:

P es la matriz de variancia y covariancias fenotípicas, de orden $n \times n$, siendo n el número de caracteres incluidos en el índice.

$$P = \begin{pmatrix} {}^P\sigma_1^2 & {}^P\sigma_{12} & \dots & {}^P\sigma_{1n} \\ {}^P\sigma_{12} & {}^P\sigma_2^2 & \dots & {}^P\sigma_{2n} \\ \dots & \dots & \dots & \dots \\ {}^P\sigma_{1n} & {}^P\sigma_{2n} & \dots & {}^P\sigma_n^2 \end{pmatrix}$$

Observemos que P posee una columna por cada carácter, es una matriz cuadrada, simétrica, cuyos elementos ubicados en la diagonal principal son las variancias fenotípicas y los restantes son las covariancias fenotípicas. La matriz P se calcula a partir de las observaciones experimentales.

La matriz G es la matriz de variancias y covariancias genotípicas, de orden $m \times m$, formada por las variancias y covariancias genéticas de los m caracteres que se quieren mejorar.

$$G = \begin{pmatrix} \sigma_1^2 & \sigma_{12}^G & \dots & \sigma_{1m}^G \\ \sigma_{12}^G & \sigma_{G2}^2 & \dots & \sigma_{2m}^G \\ \dots & \dots & \dots & \dots \\ \sigma_{1m}^G & \sigma_{Gm}^G & \dots & \sigma_m^2 \end{pmatrix}$$

Podemos definir una tercera matriz, C, tal que:

$$C = \begin{pmatrix} \sigma_{p1G1} & \sigma_{p1G2} & \dots & \sigma_{p1Gm} \\ \sigma_{p2G1} & \sigma_{p2G2} & \dots & \sigma_{p2Gm} \\ \dots & \dots & \dots & \dots \\ \sigma_{pnG1} & \sigma_{pnG2} & \dots & \sigma_{pnGm} \end{pmatrix}$$

Vemos que C es la matriz de covariancias entre los valores fenotípicos de cada uno de los n caracteres y los respectivos valores genotípicos de los m caracteres. En el caso de que $n=m$ ocurre que $C=G$.

Debe definirse además una serie de vectores:

$a' = (a_1 \quad a_2 \quad a_3 \quad \dots \quad a_m)$ es el vector de pesos económicos,

$b' = (b_1 \quad b_2 \quad b_3 \quad \dots \quad b_n)$ es el vector de pesos del índice,

$p' = (p_1 \quad p_2 \quad p_3 \quad \dots \quad p_n)$ es el vector de valores fenotípicos (observaciones) de cada individuo, familia, o unidad de selección,

$g' = (g_1 \quad g_2 \quad g_3 \quad \dots \quad g_m)$ es el vector de valores genotípicos (no observables) de cada individuo, familia o unidad de selección.

Considerando que los valores genotípicos (g) y fenotípicos (p) se determinan como desvío de sus respectivas medias, podemos establecer la relación entre las matrices y vectores indicados más arriba del siguiente modo:

$$\varepsilon (pp') = P$$

$$\varepsilon (gg') = G$$

$$\varepsilon (pg') = C$$

$$I = b'p$$

$$H = a'g$$

Las variancias del índice (I) y del genotipo agregado (H) y la covariancia entre H e I expresadas en términos matriciales son:

$$V_I = V_{(b'p)} = \text{COV}_{(b'p,b'p)}$$

Recordando que $pp' = P$:

$$V_I = \sigma_I^2 = b'Pb, \quad [12.13]$$

$$V_H = V_{(a'g)} = \text{COV}_{(a'g,a'g)}$$

Asimismo $gg' = G$:

$$V_H = \sigma_H^2 = a'Ga, \quad [12.14]$$

$$\text{COV}_{HI} = \text{COV}_{(b'p,a'g)}$$

Y $pg = C$:

$$\text{COV}_{HI} = \sigma_{HI} = b'Ca, \quad (\text{si } n \neq m) \quad [12.15]$$

La covariancia entre los valores fenotípicos para el carácter i -ésimo y los valores genotípicos para otro carácter j -ésimo puede expresarse como

$$\text{COV}(P_i, G_j) = \text{COV}(G_i + E_i, G_j) = \text{COV}(G_i, G_j) + \text{COV}(E_i, G_j)$$

El último término es cero por cuanto los desvíos ambientales del i -ésimo carácter son independientes de los valores genotípicos para el j -ésimo carácter. Por lo tanto:

$$\text{COV}(P_i, G_j) = \text{COV}(G_i, G_j)$$

Así

$$\text{COV}_{HI} = \sigma_{HI} = b'Ga, \quad [12.16]$$

Recordemos que si se dispone de la misma cantidad de información para los caracteres incorporados en el índice:

$$V_I = b'Pb = b'Ga = \text{COV}_{HI}$$

El coeficiente de regresión entre el índice y el genotipo agregado es:

$$b_{HI} = \sigma_{HI} / \sigma_I^2 = b'Ca / b'Pb$$

o bien

$$b_{HI} = \sigma_{HI} / \sigma_I^2 = b'Ga / b'Pb$$

La eficiencia del índice como instrumento de selección será mayor cuanto mayor sea la correlación entre el índice fenotípico (I) y el índice genotípico o genotipo agregado (H), que en términos matriciales será:

$$r_{HI} = \text{COV}_{HI} / (V_I V_H)^{0.5} = b'Ca / [(a'Ga)(b'Pb)]^{0.5}, \quad [12.17]$$

o bien

$$r_{HI} = \text{COV}_{HI} / (V_I V_H)^{0.5} = b'Ga / [(a'Ga)(b'Pb)]^{0.5} = [b'Gb / a'Ga]^{0.5}, \quad [12.18]$$

La covariancia entre el valor genotípico para el carácter i-ésimo que se pretende mejorar y el índice es:

$\text{COV}_{(Gi,I)} = \sigma_{(Gi,I)} = b'c_i$, donde c_i es la i-ésima columna correspondiente al carácter a mejorar dentro de la matriz C, o bien dentro de la matriz G.

Puede demostrarse (Wricke y Weber, 1986) que la correlación alcanzará su máximo cuando:

$$Pb = Ca; \text{ o bien } Pb = Ga \quad [12.19]$$

Por lo tanto el vector de pesos del índice debe ser:

$$b = P^{-1}Ca = P^{-1}Ga, \quad [12.20]$$

12.3.5. Pesos del índice

Como se mencionó más arriba, los pesos deben ser tales que maximicen la correlación entre el índice y el genotipo agregado (r_{HI}), o bien que minimicen el desvío estándar de la diferencia entre el índice y el genotipo agregado ($\sigma_{(I-H)}$).

Recordando que $Pb = Ga$

$$b = P^{-1}Ga$$

12.3.6. Respuesta a la selección basada en índices de selección

Dado el índice $I = b'p$, recordemos las ecuaciones pertinentes:

$$\begin{aligned} \sigma_{IH} &= b'Ga = \sigma_I^2 \\ \sigma_I^2 &= b'Pb \\ \sigma_H^2 &= a'Ga \\ b_{HI} &= \sigma_{HI} / \sigma_I^2, \end{aligned}$$

$$\text{COV}(G,I) = \sigma_{GiI} = b'c_i$$

$$r_{IH} = b'Ga / [(b'Pb) (a'Ga)]^{0,5}$$

Entonces la respuesta al índice de selección aplicando una intensidad de selección i será:

$$R_i = i r_{IH} \sigma_H = i \sigma_{HiI} / (\sigma_I^2 \sigma_H^2)^{0,5} \sigma_H = i [\sigma_I^2 / (\sigma_I \sigma_H)] \sigma_H = i \sigma_I = i (b'Pb)^{0,5}$$

El vector Δ o de respuesta para los caracteres contemplados en el índice agregado H basado en la selección por el índice I tiene como elementos, precisamente, a las respuestas a la selección para cada uno de los m caracteres individuales contemplados en el índice genotípico:

$$\Delta = \begin{pmatrix} \Delta_{G1} \\ \Delta_{G2} \\ \dots \\ \Delta_{Gn} \end{pmatrix} = iGb / (b'Pb)^{0,5}$$

La respuesta esperada a la selección para cualquier carácter i -ésimo incluido en el índice será:

$$\Delta_{Gi} = i b_{GiI} \sigma_i = i r_{GiI} \sigma_i$$

O bien:

$$R_i = i \sigma_{(Gi,I)} / (\sigma_I^2 \sigma_i^2)^{0,5}$$

= $i b'c_i / (b'Pb \sigma_i^2)^{0,5}$, siendo c_i la i -ésima columna de la matriz C .

La respuesta en el mérito agregado es la sumatoria de las respuestas para cada uno de los m caracteres incluidos en el índice ponderados por los respectivos pesos económicos.

$$R_H = \Delta_H = \sum a_i \Delta_{Gi} = a' \Delta \quad [12.21]$$

12.3.7. Índice óptimo e Índice estimado

Cuando en condiciones ideales se conocen los elementos de la matriz G , es decir, las variancias y covariancias genéticas paramétricas de los n caracteres para seleccionar y se las utiliza para definir un índice, este se denomina índice óptimo o de Smith (1936) y Hazel (1943). La utilización de los índices de Smith-Hazel requiere además que las medias de las unidades de selección sean conocidas y que la cantidad de información disponible sea igual para todas ellas. Por el contrario el término índice estimado se construye basado en estimaciones de la matriz de variancias y covariancias genéticas.

Recordemos la ecuación del índice:

$$I = \sum_{i=1}^n b_i p_i$$

Smith demostró que los coeficientes del índice óptimo que maximizan r_{HI} se obtienen como solución del siguiente sistema de ecuaciones:

$$P\mathbf{b} = G\mathbf{a} \quad [12.22]$$

de modo que

$$\mathbf{b} = P^{-1}G\mathbf{a} \quad [12.23]$$

En el caso de un índice con $n = 2$ caracteres los pesos del índice óptimo son:

$$b_1 = {}^P\sigma_2^2 \sigma_{P1H} - {}^P\sigma_{12} \sigma_{P2,H}$$

$$b_2 = {}^P\sigma_1^2 \sigma_{P2,H} - {}^P\sigma_{12} \sigma_{P1,H}$$

Suponiendo que se considerara el mejoramiento de tres caracteres, denominados sucesivamente 1, 2 y 3, las ecuaciones del índice óptimo de Smith serán:

$$\left. \begin{aligned} b_1 {}^P\sigma_1^2 + b_2 {}^P\sigma_{12} + b_3 {}^P\sigma_{13} &= a_1 {}^G\sigma_1^2 + a_2 {}^G\sigma_{12} + a_3 {}^G\sigma_{13} \\ b_1 {}^P\sigma_{12} + b_2 {}^P\sigma_2^2 + b_3 {}^P\sigma_{23} &= a_1 {}^G\sigma_{12} + a_2 {}^G\sigma_2^2 + a_3 {}^G\sigma_{23} \\ b_1 {}^P\sigma_{13} + b_2 {}^P\sigma_{23} + b_3 {}^P\sigma_3^2 &= a_1 {}^G\sigma_{13} + a_2 {}^G\sigma_{23} + a_3 {}^G\sigma_3^2 \end{aligned} \right\} [12.24]$$

Hazel, a partir del desarrollo de Smith, modificó las ecuaciones del índice partiendo de la siguiente ecuación de respuesta:

$$\Delta_H = i b_{HI} \sigma_I = i r_{HI} \sigma_H$$

Como se explicó anteriormente, debido a que σ_H es una constante o parámetro poblacional, la única manera de maximizar la respuesta a la selección es aumentar el producto $b_{HI} \sigma_I$ o bien la correlación entre el índice y el genotipo agregado (r_{HI}). A partir de las ecuaciones de Smith [12.24] si se dividen ambos miembros de la igualdad por el producto $\sigma_H \sigma_{Pi}$ obtendremos:

$$b_1 {}^P\sigma_1 / \sigma_H + b_2 {}^P r_{12} {}^P\sigma_2 / \sigma_H + b_3 {}^P r_{13} {}^P\sigma_3 / \sigma_H = r_{P1H}$$

$$b_1 {}^P r_{12} {}^P\sigma_1 / \sigma_H + b_2 {}^P\sigma_2 / \sigma_H + b_3 {}^P r_{23} {}^P\sigma_3 / \sigma_H = r_{P2H}$$

$$b_1 {}^P r_{13} {}^P\sigma_1 / \sigma_H + b_2 {}^P r_{23} {}^P\sigma_2 / \sigma_H + b_3 {}^P\sigma_3 / \sigma_H = r_{P3H}$$

Suponiendo que se mejoren simultáneamente dos caracteres, el sistema de ecuaciones del índice puede expresarse en términos de variancias y covariancias genéticas y fenotípicas, pero también en términos de correlaciones.

En el primer caso:

$$b_1^P \sigma_1^2 + b_2^P \sigma_{12} = a_1^G \sigma_1^2 + a_2^G \sigma_{12}$$

$$b_1^P \sigma_{12} + b_2^P \sigma_2^2 = a_1^G \sigma_{12} + a_2^G \sigma_2^2$$

En el segundo caso:

$$b'_1 + b'_2 r_{12} = a'_1 h_1 + a'_2 h_1 r_{12}$$

$$b'_1 r_{12} + b'_2 = a'_1 h_2 r_{12} + a'_2 h_2$$

donde:

$$b'_i = (r_{iH}^P / \sigma_H) b_i$$

$$a'_i = (r_{iH}^G / \sigma_H) a_i$$

12.3.7.1. Empleo de caracteres secundarios o variables indicadoras

Como se mencionó más arriba, si los pesos económicos se igualan a cero en algunos de los caracteres incluidos en el índice, estos actuarán como caracteres secundarios o variables indicadoras. Es decir, que la información fenotípica para esos caracteres debe estar disponible, pero no es objetivo de la selección obtener una respuesta en ellos. La utilidad de los caracteres secundarios reside en que pueden contribuir a mejorar la respuesta en el genotipo agregado porque agregan información.

El sistema de ecuaciones conducente a la estimación de los pesos del índice debe entonces modificarse. Los elementos de la derecha del sistema de ecuaciones ubicados en las columnas de los caracteres secundarios se eliminan y por lo tanto no será necesario disponer de estimaciones de sus variancias y covariancias genotípicas.

Los términos a_i son los pesos o valores económicos relativos. En el caso de variables indicadoras $a_i = 0$ puesto que no es el objetivo mejorar esos caracteres. Suponiendo que se miden cinco caracteres, designados 1, 2, 3, 4, 5 y solo se quisiera mejorar el carácter 1 (con una ponderación de 1), el vector a' sería:

$$a' = (1 \quad 0 \quad 0 \quad 0 \quad 0)$$

Si, en cambio, se deseara mejorar el carácter 3:

$$a' = (0 \quad 0 \quad 1 \quad 0 \quad 0)$$

Asumamos que disponemos de información fenotípica para tres caracteres y que se desea mejorar solamente la media genotípica de la población para el carácter 1. En tal caso podemos escribir las ecuaciones del índice de Smith de la siguiente manera:

$$\begin{aligned} b_1^P \sigma_1^2 + b_2^P \sigma_{12} + b_3^P \sigma_{13} &= G\sigma_1^2 \\ b_1^P \sigma_{12} + b_2^P \sigma_2^2 + b_3^P \sigma_{23} &= G\sigma_{12} \\ b_1^P \sigma_{13} + b_2^P \sigma_{23} + b_3^P \sigma_3^2 &= G\sigma_{13} \end{aligned}$$

En forma matricial esta expresión puede reescribirse como:

$$Pb = g$$

donde g es la columna de variancias y covariancias genotípicas del único carácter para mejorar.

Asumamos que se van a emplear dos caracteres secundarios que presentan mayor heredabilidad que el carácter primario para mejorar y alta correlación genotípica. En este caso:

A = Carácter primario que interesa mejorar

B, C = Caracteres secundarios

A, B, C = Caracteres para incluir en el índice

Será necesario tener ciertas estimaciones de las variancias fenotípicas y genéticas ($^P\sigma_i^2, ^G\sigma_i^2$) de los caracteres involucrados y de las covariancias fenotípicas y genéticas ($^P\sigma_{ij}, ^G\sigma_{ij}$). Particularmente en este caso:

$$\begin{array}{cccc} ^G\sigma_A^2 & ^G\sigma_{AB} & ^P\sigma_A^2 & ^P\sigma_{AB} \\ & ^G\sigma_{AC} & ^P\sigma_B^2 & ^P\sigma_{AC} \\ & & ^P\sigma_C^2 & ^P\sigma_{BC} \end{array}$$

Las ecuaciones del índice serán:

$$\begin{aligned} b_1^P \sigma_A^2 + b_2^P \sigma_{AB} + b_3^P \sigma_{AC} &= G\sigma_A^2 \\ b_1^P \sigma_{AB} + b_2^P \sigma_B^2 + b_3^P \sigma_{BC} &= G\sigma_{AB} \\ b_1^P \sigma_{AC} + b_2^P \sigma_{BC} + b_3^P \sigma_C^2 &= G\sigma_{AC} \end{aligned}$$

Puede observarse que hay tantas ecuaciones como caracteres medidos, la primera en relación con A , la segunda con B y la tercera con C . En cada ecuación aparece el subcripto correspondiente en cada término. La primera

columna de la izquierda lleva el subscrito A, la segunda el B, la tercera el C. Las variancias y covariancias del miembro de la izquierda son fenotípicas.

Cuando fila y columna coinciden con el subscrito se colocan variancias, caso contrario covariancias. El término de la derecha del sistema de ecuaciones tiene una sola columna y comparten el subíndice del carácter para mejorar. Solo aparecen parámetros genéticos. Si el carácter para mejorar no fuera incluido en el índice (medido), el término de la derecha solo tendría covariancias.

Las ecuaciones del índice de Hazel deberían reescribirse como (Baker, 1986):

$$b_1^P \sigma_1^P / \sigma_1^G + b_2^P r_{12}^P \sigma_2^P / \sigma_1^G + b_3^P r_{13}^P \sigma_3^P / \sigma_1^G = h_1$$

$$b_1^P r_{12}^P \sigma_1^P / \sigma_1^G + b_2^P \sigma_2^P / \sigma_1^G + b_3^P r_{23}^P \sigma_3^P / \sigma_1^G = h_2$$

$$b_1^P r_{13}^P \sigma_1^P / \sigma_1^G + b_2^P r_{23}^P \sigma_2^P / \sigma_1^G + b_3^P \sigma_3^P / \sigma_1^G = h_3$$

La respuesta esperada a la selección para mejorar el carácter A utilizando como variables del índice al mismo carácter y a los caracteres B y C como secundarios será:

$$R_1 = i^G \sigma_1^2 r_{G11}$$

Donde σ_1^2 es la variancia genotípica del carácter para mejorar y r_{G11} es la correlación genotípica entre el carácter para mejorar (A) y el índice.

La respuesta a la selección para el carácter A por sí solo (sin emplear variables indicadoras) hubiese sido:

$$R_1 = i^G \sigma_1^2 r_{G1P1}$$

Donde σ_1^2 es la variancia genotípica del carácter para mejorar y r_{G1P1} es la correlación entre los valores genotípicos y fenotípicos del carácter para mejorar. Esta última correlación es por definición igual a la raíz cuadrada de la heredabilidad del carácter (h_1).

12.3.7.2. Índice óptimo de Henderson

El empleo de este índice requiere decidir previamente, para un grupo de n caracteres, cuáles serán los m caracteres primarios y cuáles los caracteres secundarios (n-m). Posteriormente se calculan índices de selección separadamente para cada uno de los caracteres, es decir, para los m caracteres primarios (Henderson, 1963). Para ello no es necesario determinar los pesos económicos ya que cada carácter primario correspondiente a cada índice por separado recibe un peso económico implícito igual a uno.



Los m índices por separado serán:

$$\begin{aligned}
 I_1 &= b_{11} p_1 + b_{12} p_2 + \dots + b_{1n} p_n \\
 I_2 &= b_{21} p_1 + b_{22} p_2 + \dots + b_{2n} p_n \\
 I_3 &= b_{31} p_1 + b_{32} p_2 + \dots + b_{3n} p_n \\
 &\dots\dots\dots \\
 I_m &= b_{m1} p_1 + b_{m2} p_2 + \dots + b_{mn} p_n
 \end{aligned}$$

Los b_{ij} correspondientes al o los caracteres que no contribuyen al mejoramiento del carácter primario serán iguales a cero.

En la siguiente etapa se construye el índice final (I') a partir de la expresión:

$$\begin{aligned}
 I' &= a_1 I_1 + a_2 I_2 + \dots + a_m I_m \\
 &= (a_1 b_{11} p_1 + a_1 b_{12} p_2 + \dots + a_1 b_{1n} p_n) \\
 &\quad + (a_2 b_{21} p_1 + a_2 b_{22} p_2 + \dots + a_2 b_{2n} p_n) \\
 &\quad + \dots + (a_m b_{m1} p_1 + a_m b_{m2} p_2 + \dots + a_m b_{mn} p_n) \\
 &= (a_1 b_{11} + a_2 b_{21} + \dots + a_m b_{m1}) p_1 \\
 &\quad + (a_1 b_{12} + a_2 b_{22} + \dots + a_m b_{m2}) p_2 \\
 &\quad + \dots + (a_1 b_{1n} + a_2 b_{2n} + \dots + a_m b_{mn}) p_n
 \end{aligned}$$

La ventaja del índice de Henderson reside en la facilidad con que puede calcularse un nuevo índice I' ante modificaciones de los pesos económicos de los caracteres primarios (objetivos del mejoramiento) sin necesidad de recalcular los índices individuales.

Asumamos que desea diseñarse un índice de Henderson a partir de un grupo de n=5 caracteres (A, B, C, D y E), que se desean mejorar los caracteres A y D, que los tres primeros caracteres no son útiles para mejorar el carácter D y que tanto D como E no sirven para mejorar el carácter A. Esta información la utilizaremos para definir los pesos de los índices individuales que serán iguales a cero.

El índice apropiado para mejorar el carácter primario A será de la forma:

$$I_A = b_A p_A + b_B p_B + b_C p_C$$

Nótese que con base en el supuesto de que ni D ni E resultan útiles para mejorar el carácter primario A, sus pesos del índice b serán iguales a cero y los términos correspondientes por lo tanto desaparecen de la expresión.

Asumamos un índice para mejorar el carácter D como:

$$I_D = b_D p_D - b_E p_E$$

Si los caracteres A y D tuvieran pesos económicos 1 y 0,3, respectivamente, el mejor índice para mejorar A y D sería:

$$I' = (1)I_A + (0,3) I_D = (1)(b_A p_A + b_B p_B + b_C p_C) + (0,3)(b_D p_D - b_E p_E)$$

Si en cambio los caracteres A y D tuvieran pesos económicos 1 y 1,5 el mejor índice para mejorar A y D sería:

$$I' = (1)(b_A p_A + b_B p_B + b_C p_C) + (1,5) (b_D p_D - b_E p_E)$$

12.3.8. Índice base o de Williams

Las estimaciones de variancias y covariancias necesarias para calcular el índice de selección tienen asociados errores de estimación. En circunstancias en las que los errores son importantes, puede resultar más conveniente emplear el denominado índice base (Williams, 1962). El índice base es aplicable cuando los caracteres genotípicos que interesan mejorar coinciden con los caracteres fenotípicos que se miden. En otras palabras, todos los caracteres son primarios.

En el índice base de Williams se utilizan como pesos del índice a los pesos económicos, es decir que $b_i = a_i$.

El índice base en forma matricial toma la forma:

$$I_B = a'P$$

Y por lo tanto:

$$\sigma_I^2 = a'Pa$$

$$\sigma_H^2 = a'Ga$$

$$r_{IH} = (a'Ga/a'Pa)^{0,5}$$

12.3.8.1. Índice base modificado

Para las situaciones en que los caracteres para seleccionar poseen pesos económicos iguales o muy similares, Smith *et al.* (1981) propusieron el empleo de un índice base modificado en el que los pesos del índice son las heredabilidades de cada carácter:

$$I = a_1 h^2_1 p_1 + a_2 h^2_2 p_2 + \dots + a_n h^2_n p_n$$

O bien:

$$I = \sum a_i h^2_i P_i$$

Matricialmente:

$$I_{BM} = (ah^2)'P$$

En este caso los pesos económicos son positivos o negativos, según que el carácter contribuya favorable o desfavorablemente al valor genotípico agregado. Smith *et al.* (1981) sostienen que este índice resulta muy eficiente cuando los caracteres incorporados al índice no están correlacionados, o sus coeficientes de correlación en valores absolutos resultan ser menores a 0,30.

Si los pesos económicos fueran iguales, el índice base modificado expresado en forma matricial es:

$$I_{BM} = (h^2)'P$$

La eficiencia del índice base es similar a la del índice óptimo cuando los caracteres del índice están escasamente correlacionados y la magnitud relativa de sus variancias resulta ser una buena medida de su importancia relativa (Baker, 1986).

12.3.9. Índices restringidos

En un índice clásico las características de los cambios por selección que pueden ocurrir en la media de los caracteres componentes de H responderán a la matriz de correlaciones genéticas que los vincula. Ocasionalmente puede suceder que la aplicación de un índice de selección clásico (I) no lleve a lograr mejoras en la magnitud y sentido deseado para todos los caracteres incorporados en el mérito agregado (H). Ello puede ocurrir cuando un carácter tiene baja heredabilidad y existen diferencias importantes entre las matrices de variancias y covariancias fenotípicas y genotípicas (Wricke y Weber, 1986).

El objetivo del fitomejorador puede ser mejorar varios caracteres que componen H, pero al mismo tiempo tratar de evitar cambios indeseables en otros. Para evitar situaciones como las mencionadas más arriba, pueden utilizarse índices restringidos (Kempthorne y Nordskog, 1959; Tallis, 1962; Cunningham *et al.*, 1970) e índices de ganancias deseadas (Pesek y Baker, 1969; Harville, 1975; Tai, 1977). En cierto modo el índice restringido de Kempthorne y Nordskog constituye un caso especial del índice de ganancias deseadas, en el que se busca lograr una respuesta nula a la selección por el índice para uno o más caracteres.

Los índices restringidos admiten la incorporación de caracteres secundarios y requieren disponer de estimaciones confiables de las variancias y covariancias genéticas. La imposición de restricciones en la respuesta para ciertos caracteres reduce la correlación entre el índice de selección y el mérito agregado (r_{HI}) y consecuentemente con ello la respuesta esperada a la selección disminuye (Baker, 1986).

12.3.9.1. Índices restringidos de Kempthorne y Nordskog (1959), Cunningham et al. (1970) y Tallis (1962)

Esencialmente los pesos del índice (b_i) en los índices restringidos se eligen de modo tal que algunas de las covariancias entre el índice I y los valores de mejora del carácter que se restringe (g_i) se igualan a cero. El objetivo es maximizar el genotipo agregado restringiendo el cambio en algunos caracteres ($R_{Gi}=0$).

En el caso de que se restringiera (anulara) la respuesta al carácter 1 empleando un índice fenotípico que considera n caracteres:

$$\sigma_{I_g1} = b_1\sigma_{p1g1} + b_2\sigma_{p2g1} + \dots + b_n\sigma_{png1} = (c_1)'b = 0$$

donde c_1 es la primera columna de la matriz C (o G), es decir, la que corresponde al carácter 1.

La relación entre un índice restringido (I_R) y no restringido (I) en términos matriciales adopta la siguiente expresión:

$$I_R = IM$$

donde M es una matriz modificatoria.

Cunningham et al. (1970), por su parte, propusieron un índice restringido a partir de:

$$P^*b^* = G^*a$$

$$\left| \begin{array}{cc|c|c} P & G_i & b & G \\ \hline G_i' & 0 & b_d & 0 \end{array} \right| = \quad (a)$$

Donde G_i es la columna i -ésima de G correspondiente al carácter i -ésimo que se restringe.

$$b = [I - P^{-1}G_i(G_i' P^{-1}G_i)^{-1} G_i']P^{-1}G a$$

$$b_d = (G_i' P^{-1}G_i)^{-1} G_i' P^{-1}G a$$

El valor de b_d resulta ser el peso económico con signo negativo que sería necesario emplear para producir una respuesta nula en el carácter restringido si se hubiera aplicado un índice clásico. Se trata de una variable *dummy* que no se utiliza en la ecuación del índice.

Los desarrollos de Cunnigham *et al.* y de Kempthorne y Nordkog son equivalentes y conducen por lo tanto a la estimación de pesos del índice idénticos.

Tallis (1962) propuso un índice restringido, pero con una restricción menos estricta. En este caso $R_{Gi} = k_i$, donde k_i es una constante diferente a cero. Cuando $k_i = 0$, el índice restringido de Tallis y el de Kempthorne y Nordskog coinciden.

12.3.9.2. Índice de ganancias o respuestas deseadas (Pesek y Baker, 1969; Harville, 1975; Tai, 1977)

El índice de ganancias deseadas (Pesek y Baker, 1969) no admite el empleo de caracteres secundarios para mejorar la respuesta a la selección en los caracteres principales. En otras palabras, los caracteres que se incorporan en el índice coinciden con los caracteres que se quieren mejorar ($n=m$). Este índice es similar al de Tallis (1962), pero las restricciones se aplican a todos los caracteres del índice.

Cuando es difícil definir los pesos económicos a_i conviene expresarlos como objetivos específicos de selección, en términos de valores óptimos para cada carácter. El índice de ganancias deseadas no requiere establecer pesos económicos, sino indicar en su lugar el genotipo deseado, que se expresa en el vector de ganancias o respuestas deseadas (R_{gd}). Así por ejemplo, si se desean mejorar tres caracteres (A, B y C) y se desea lograr una respuesta tres veces mayor en el carácter B que en el A y del doble para el carácter C que en el A, el vector sería el siguiente:

$$R_{gd} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$$

El índice de ganancias deseadas toma la forma:

$$I = b'P,$$

y los pesos del índice se determinan como:

$$b' = G^{-1}R_{gd}$$

En todo caso los pesos económicos pueden calcularse *a posteriori*, una vez conocidos los pesos del índice, mediante la ecuación:

$$a = G^{-1}Pb$$

Los antecedentes de utilización del índice de ganancias deseadas sugieren que para que este índice resulte eficiente esas ganancias deberían ser proporcionales a la magnitud de los desvíos estándar de los caracteres que interesa seleccionar.

El índice de ganancias deseadas de Harville (1975) es una generalización del índice de Pesek y Baker (1969), pero a diferencia de este último permite incorporar en el índice caracteres o variables secundarias. Si no se utilizan variables secundarias, las estimaciones de los pesos del índice de Pesek y Baker y el de Harville resultan iguales.

Los índices restringidos de Pesek y Baker, Tai y Harville son equivalentes al índice de Tallis cuando las restricciones se aplican a todos los caracteres del índice. En circunstancias en que solo algunos de los caracteres se restringen, el índice de Tallis resultaría menos eficiente que los de Harville y Tai (Baker, 1986).

12.3.10. Índice multiplicativo

Elston (1963) propuso un índice multiplicativo de los valores fenotípicos. Requiere la transformación previa de los datos observados de cada carácter, x_i , por el valor mínimo k_i establecido por el fitomejorador para seleccionar. De tal manera:

$$I_e = (x_1 - k_1)(x_2 - k_2)\dots(x_n - k_n) = \prod(x_i - k_i)$$

Una vez calculados los valores del índice para las unidades de selección se ordenan de mayor a menor. Los valores fenotípicos de cada carácter se transforman para que todos ellos tengan variancia igual a 1.

El índice multiplicativo de Elston posee la ventaja de no requerir estimaciones de los parámetros genéticos o fenotípicos y resulta apropiado cuando es difícil estimar los pesos económicos. Implícitamente adjudica a cada carácter un peso económico similar, lo cual significa una ventaja cuando no hay manera objetiva de determinar su importancia relativa.

Cuando los histogramas de los caracteres observados son muy diferentes, Elston (1963) propuso aplicar una transformación, a fin de que los histogramas de todas las variables sean parecidos en posición y que todos los valores sean positivos:

$$x'_{ij} = \log(x_{ij} - k_i)$$

Una vez definidos los valores x'_{ij} , debe definirse a k'_i basada en la expresión $k'_i = n (\min. x_{ij} - \max x_{ij}) / (n-1)$, donde n es el número de unidades de selección. De tal manera el índice multiplicativo se construye como:

$$I'_e = (x'_1 - k'_1)(x'_2 - k'_2) \dots (x'_n - k'_n) = \prod (x'_i - k'_i)$$

Baker sostiene que este índice sería equivalente a un índice base, donde los pesos de cada carácter son:

$$b_i = 1/p\sigma_i,$$

donde $p\sigma_i$ es el desvío estándar fenotípico del i -ésimo carácter incorporado en el índice fenotípico.

Esta opción resulta apropiada cuando los caracteres tienen importancia similar para los objetivos del fitomejorador. Si los caracteres tuvieran diferente importancia, sería más conveniente utilizar el índice base de Williams (Baker, 1986).

La respuesta en el genotipo agregado empleando el índice de Elston es equivalente a la obtenible con el índice de Smith y Hazel donde $b_i = 1/p\sigma_i$ (Crosbie *et al.*, 1980).

12.3.11. Índice de suma de *rankings*

Fue propuesto por Mulamba y Mock (1978) y el índice se construye a partir de:

$$I = \sum \text{ranking } (p_i)$$

Donde p_i hace referencia al valor observado de cada genotipo para el i -ésimo carácter considerado en el índice.

Al igual que para el índice multiplicativo, la respuesta esperada sería parecida a la obtenible con un índice de Smith y Hazel, donde los coeficientes del índice son los recíprocos de los desvíos estándar fenotípicos.

12.3.12. Otros índices no lineales

Kempthorne y Nordskog (1959) postularon un índice de selección en el que el genotipo agregado es una función cuadrática del valor genotípico. Otro tipo de índices resultaría como producto o cociente entre variables (en escala logarítmica son lineales).

El índice de Schwarzbach (1972), citado por Wricke y Weber (1986), considera $m=n$ y un genotipo ideal definido por el fitomejorador. Para cada genotipo se miden las distancias fenotípicas respecto a ese genotipo ideal. Pueden emplearse medidas de distancias euclidianas, de Mahalanobis, etc.

A modo de ejemplo:

$$D_{k\text{-ésimo fenotipo}} = [\sum a_i (p_{ki} - g_i)^2]^{0,5}$$

donde:

D_k es el índice de distancia euclidiana entre el individuo o familia k-ésima respecto del genotipo ideal u objetivo del fitomejorador, medido sobre i caracteres que se incluyen en el índice;

a_i es el peso económico del carácter;

p_{ik} es la medida fenotípica observada del i-ésimo en el individuo o familia k-ésima

y g_i es el valor genotípico ideal para el carácter j.

12.4. Comparación de la eficiencia de diferentes estrategias de selección para caracteres múltiples

Las eficiencias relativas de estrategias selectivas alternativas dependerán de si los caracteres bajo selección están o no genéticamente correlacionados.

Hazel y Lush (1942), citados por Baker (1986) compararon la respuesta esperada a la selección de m caracteres independientes y con distribución normal, en términos de sus heredabilidades (h^2), pesos económicos (a_m) y variancias fenotípicas ($^P\sigma_m$) para tres modalidades de selección. La respuesta esperada empleando un índice de selección será:

$$R_I = i (a_1^2 h_1^4 {}^P\sigma_1^2 + a_2^2 h_2^4 {}^P\sigma_2^2 + \dots + a_m^2 h_m^4 {}^P\sigma_m^2)^{0,5}$$

La respuesta a la selección por niveles de corte independientes:

$$R_{NCl} = i_1 a_1 h_1^2 {}^P\sigma_1 + i_2 a_2 h_2^2 {}^P\sigma_2 + \dots + i_m a_m h_m^2 {}^P\sigma_m$$

y la respuesta a la selección en tándem (para cada etapa secuencial):

$$R_{ST} = i {}^P\sigma h^2$$

Asumiendo que los productos entre los pesos económicos, la heredabilidad y el desvío estándar fenotípico fueran los mismos para cualquier carácter incluido en el índice, es decir que:

$$a_i h_i^2 {}^P\sigma_i = c,$$

donde c es una constante, se cumple que:

$$R_I = (m)^{0,5} i a h^2 P \sigma$$

$$R_{NCl} = m k' a h^2 P \sigma$$

$$R_{ST} = i a h^2 \sigma_p,$$

donde k' depende de la fracción de la población que se selecciona aplicando cada uno de los niveles independientes de corte.

Bajo estas condiciones concluyeron que aplicando la misma intensidad de selección la respuesta al índice de selección será mayor a la de la selección en tándem y tanto mayor cuanto más caracteres sean considerados. La eficiencia de la selección por niveles de corte independiente se puede aproximar a la del índice de selección cuanto menos caracteres se seleccionen y se usen mayores intensidades de selección. La ventaja de R_{NCl} respecto de R_{ST} aumenta con el número de caracteres y cuanto mayor es la intensidad de selección.

Young (1961) concluyó que, por una parte, R_I es mayor que R_{ST} cuanto mayor es m y cuanto más similares sean los pesos económicos. Por otra parte, R_I es mejor que R_{NCl} cuando los pesos económicos son similares y las intensidades de selección son bajas o intermedias.

Cuando los m caracteres bajo selección están correlacionados, las eficiencias relativas de R_I , R_{NCl} y R_{ST} cambian según cómo sean P_r , G_r y la proporción entre las covariancias de los valores genotípicos de los caracteres individuales con el genotipo agregado y el desvío estándar fenotípico de cada carácter (Young, 1961; Finney, 1962).

Cuando hay mucha diferencia entre los caracteres para la proporción ($\sigma_{Gi,H}/P\sigma_i$), la selección en tándem y en niveles de corte independientes se aproximan en eficiencia a la selección basada en índices. Si bien resulta difícil generalizar, el orden de las eficiencias expresados a través de la magnitud de las respuestas tendería a ser el siguiente:

$$R_I \geq R_{NCl} \geq R_{ST}$$

Para mejorar un carácter simple, el uso de un índice con varios caracteres, además del que se mejora, permite obtener mejores respuestas que la selección directa (Baker, 1986).

El número de caracteres incorporados en el índice determina la cantidad de parámetros genéticos que es necesario estimar y por lo tanto a mayor número de caracteres, mayor costo para aplicar el índice de selección. Una manera de determinar si conviene incorporar un carácter en el índice es

determinar su contribución relativa al mejoramiento del genotipo agregado. Cunningham (1969) propone comparar la eficiencia de índices reducidos (con $m-1$ caracteres) respecto de un índice original (con m caracteres) a partir del cociente entre sus respectivos desvíos estándares fenotípicos mediante la siguiente expresión:

$$ER = \{[b'Pb - (b_i^2/w_{ii})]/b'Pb\}^{0.5}$$

El término b_i es el que corresponde al coeficiente del índice original para el carácter que se elimina del índice reducido y w_{ii} es el elemento de la inversa de la matriz de variancias y covariancias fenotípicas que corresponde al i -ésimo carácter.

Otra medida de la eficiencia relativa es calcular la relación entre las correlaciones entre el genotipo agregado y los valores de los índices que se están comparando. Así por ejemplo, Baker (1986) llegó a concluir que la mayor ventaja del índice óptimo respecto del índice base ocurre cuando la mitad de los caracteres tiene alta heredabilidad y los restantes muy baja heredabilidad. Lo opuesto sucede cuando los caracteres incluidos en los índices no difieren tanto en su heredabilidad.

Alternativamente, pueden compararse las respuestas esperadas en el genotipo agregado empleando el índice original y los índices reducidos, o bien comparar la respuesta esperada en el i -ésimo carácter empleando el índice o haciendo selección individualmente para ese carácter.

Si las correlaciones genotípicas y fenotípicas entre los caracteres son cercanas a cero y $b = a_i \sigma_{G(i)} = k$, siempre el índice óptimo será mejor que el índice estimado. No obstante, cuando no se dispone de buenas estimaciones de las variancias y covariancias poblacionales, la eficiencia del índice estimado puede ser menor que la de un índice base. Por lo tanto convendrá utilizar el índice estimado únicamente cuando se cuenta con estimaciones confiables y realistas de los parámetros genéticos, o cuando las heredabilidades sugieren que el índice óptimo podría ser mucho mejor que el índice base.

12.5. Selección combinada basada en información molecular y fenotípica

Si los efectos de genes de interés detectados en desequilibrio de ligamiento con marcadores moleculares no explican una fracción importante de la variabilidad fenotípica del carácter, la respuesta obtenible por selección basada en su solo aumento de frecuencia puede esperarse que no sea importante. En tal circunstancia puede resultar más conveniente combinar la selección sobre las dos fuentes de variabilidad disponible:

- a. variabilidad no asociada a ningún QTL;
- b. variabilidad explicada por QTLs.

Chakraborty *et al.* (2002) y Dekkers y Van Arendok (1998) propusieron un esquema de selección asistida por marcadores para 1 o 2 QTLs con efecto suficientemente grande y fondo genético poligénico. Para ello se emplea un Índice Marcador-Fenotipo.

12.5.1. Índice Marcador-Fenotipo

Se utilizan los marcadores para mejorar la predicción del valor de mejora de un individuo y a partir de este valor se seleccionan los mejores individuos (Whittaker, 2001). En cierto modo es como si la información molecular actuara como un carácter secundario que se incluye en el índice fenotípico.

Las bases teóricas de este sistema fueron expuestas por Lande y Thompson (1990):

- A cada marcador, según su efecto se le adjudica un puntaje o *score* molecular.
- Para múltiples marcadores, el puntaje de un individuo será la sumatoria de los puntajes moleculares.
- Suponiendo que para cada individuo consideramos como dos caracteres correlacionados la información fenotípica (P) y la información molecular (M), se propone un índice lineal de la forma:

$$I = b_M M + b_P P$$

Se ha encontrado que este índice resulta más eficiente en caracteres de baja heredabilidad, pero Moreau *et al.* (1998) sostienen que en caracteres de baja heredabilidad también se reduce el poder de detección de efectos de marcadores y entonces la mejor situación para aplicarlo es cuando la heredabilidad es moderada. Si los efectos de los marcadores estuvieran sobreestimados, el índice puede poner más peso sobre la información molecular y en consecuencia afectarse (reducirse) la respuesta obtenible a partir del índice. Otra desventaja de este esquema de selección es que no se evitan los costos adicionales que se agregan con la fenotipificación, lo cual no ocurre con los esquemas de construcción de genotipos o *genotype building*, o con la selección recurrente genómica.

12.6. Aplicación del concepto de índices de selección a la selección intrapoblacional en varias etapas

En cualquier programa de mejoramiento con una disponibilidad finita de recursos, en las diferentes etapas de evaluación se emplea un número variable de repeticiones y localidades y se tiende a asignar un número mayor de recursos (combinaciones de parcelas de repeticiones y de ambientes) por genotipo en las etapas avanzadas de la selección. Es preciso proceder a un descarte de genotipos deficientes que libere recursos para evaluar con mayor heredabilidad los genotipos aparentemente más promisorios.

Asumamos que el programa dispone en total de N parcelas experimentales y que un número G de genotipos se evalúan en 1 localidad. Luego de efectuadas S etapas de selección, se llega a un número final αG de genotipos seleccionados. En cada etapa i -ésima de selección (S_i) se retiene (selecciona) una fracción α_i .

Por lo tanto la fracción final de los G genotipos iniciales será:

$$\alpha = \prod \alpha_i = \alpha_1 \alpha_2 \alpha_3 \dots \alpha_i$$

donde \prod es el operador multiplicatoria.

Asumamos también que para la i -ésima etapa de selección se dispone de N_i parcelas de evaluación. De tal modo:

$$N = \sum N_i$$

Finney (1958) propone un esquema de selección cercano al óptimo (en términos de la mayor respuesta esperable) en el cual:

$$\alpha_i = \alpha^{1/S}, \quad [12.25]$$

$$N_i = N/S, \quad [12.26]$$

A manera de ejemplo, asumiendo que hay un total de $N=30000$ parcelas experimentales que deben asignarse a 4 etapas de selección aplicadas a un número inicial de $G=1500$ genotipos y que la proporción final de genotipos seleccionados resultado de las cuatro etapas de selección fuera $\alpha=0,0007$:

Aplicando [12.26]:

$$N_i = 30000/4 = 7500 \text{ parcelas por etapa de selección, por lo tanto:}$$

$$N_1 = N_2 = N_3 = N_4 = 7500 \text{ parcelas.}$$

Y por [12.25] concluimos que:

$$\alpha_i = (0,0007)^{(1/4)} = 0,1625 \text{ es decir que}$$

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0,1625$$

La aplicación de estos criterios de Finney al ejemplo se presenta en la Tabla 12.2.

Etapas	N.º genotipos evaluados	N.º repeticiones	N.º genotipos selectos
1	$G_1=1500$	$N_1/G_1=7500/1500=5$	$\alpha_1 G_1 = (0,1625) 1500 =244$
2	$G_2=\alpha_1 G_1 =244$	$N_2/G_2=7500/244=30$	$\alpha_1 \alpha_2 G_1 = \alpha_2 G_2 = (0,1625) 244=40$
3	$G_3=\alpha_2 G_2 =40$	$N_3/G_3=7500/40=188$	$\alpha_1 \alpha_2 \alpha_3 G_1 = \alpha_3 G_3 = (0,1625) 40 = 7$
4	$G_4=\alpha_3 G_3 =7$	$N_4/G_4=7500/7=1071$	$\alpha_1 \alpha_2 \alpha_3 \alpha_4 G_1 = \alpha_4 G_4 = (0,1625) 7 =1$

Tabla 12.2

Asignación de parcelas experimentales y proporción de genotipos seleccionados en un programa de selección en cuatro etapas, partiendo de un total de 30000 parcelas disponibles, 1500 genotipos bajo selección y una proporción final de genotipos selectos de 0,0007.

La información presentada en la Tabla 12.2 es una aproximación al óptimo, puesto que no se están considerando las variancias genéticas y no genéticas (y por lo tanto la heredabilidad). El valor óptimo caerá en algún punto donde:

$$\alpha_1 < \alpha_2 \dots < \alpha_s$$

y

$$N_1 < N_2 \dots < N_s \text{ (Wricke y Weber, 1986).}$$

Cochran (1951) analizó la respuesta esperada a este tipo de selección por etapas considerando que la selección que se hace en cada una operaría como si se tratara de caracteres diferentes, tal como se realiza en la selección por niveles de corte independientes (este mismo capítulo). Considerando que se seleccionará solamente en dos etapas, finalmente la población total de genotipos podrá subdividirse en un grupo de genotipos que solo se evaluaron en la etapa 1 y no fueron seleccionados, el grupo de genotipos que se seleccionaron (retuvieron) en la etapa 1, pero no en la etapa 2 y los que fueron seleccionados en ambas etapas (Figura 12.7).

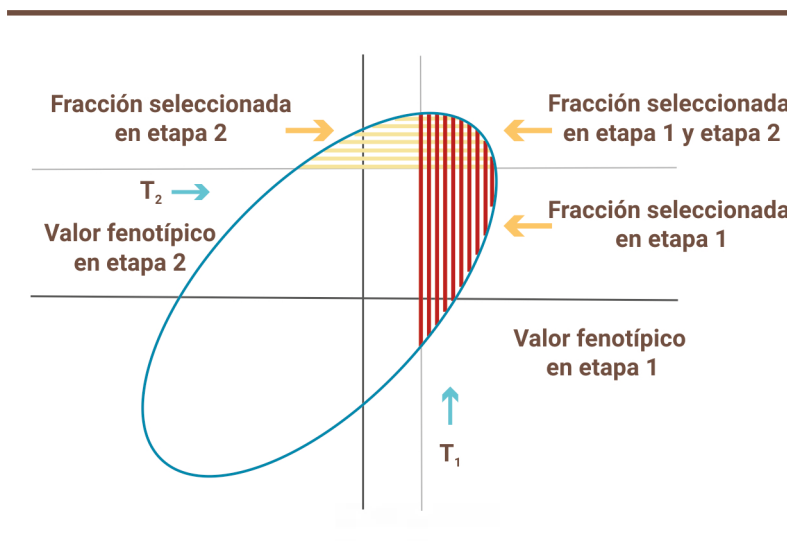


Figura 12.7. Distribución de una población de genotipos expuesta a un proceso de selección en dos etapas.

Pese a que el método de Finney no tiene en cuenta diferentes combinaciones relativas de variancias genotípicas y ambientales (heredabilidades), Cochran (1951) concluyó que para muchas de ellas la solución aproximada de Finney se acerca a la estrategia óptima (Tabla 12.3). Las columnas encabezadas con r_1 y r_2 corresponden al número de repeticiones en la primera y segunda etapas correspondientes al óptimo, redondeadas a la cifra entera más próxima).

σ_G^2 / σ_E^2	Solución óptima según Cochran					Solución óptima según Finney				
	α_1	α_2	r_1	r_2	R	α_1	α_2	r_1	r_2	R
0,1	0,22	0,28	1	6	1.050	0,25	0,25	1	4	1.039
1,0	0,21	0,30	1	4	1.772	0,25	0,25	1	4	1.770
2,0	0,20	0,31	1	5	1.874	0,25	0,25	1	4	1.869
10,0	0,15	0,42	1	7	1.956	0,25	0,25	1	4	1.948

Tabla 12.3.

Asignación de porcentajes de genotipos retenidos en cada etapa (α) y número de repeticiones (r) o parcelas de evaluación (redondeadas al número entero más próximo) y respuesta esperada (R) en un programa de selección en dos etapas, para diferentes relaciones de variancias genéticas y ambientales, de acuerdo a Cochran (1951) y siguiendo la propuesta de Finney (1958). Se asume que la fracción total seleccionada es el 6,25 % de los genotipos evaluados.

12.6.1. Respuestas con diferentes estrategias de selección por etapas

Analizaremos cuál es la respuesta a tres estrategias de selección: en una etapa, en dos etapas y en múltiples etapas.

La mayor respuesta posible a la selección (medida en desvíos estándar de la media de la población original de N genotipos) o la situación óptima, considerando que se trabajara sin error, dependerá de la fracción de genotipos seleccionados hasta la etapa final (Figura 12.8).

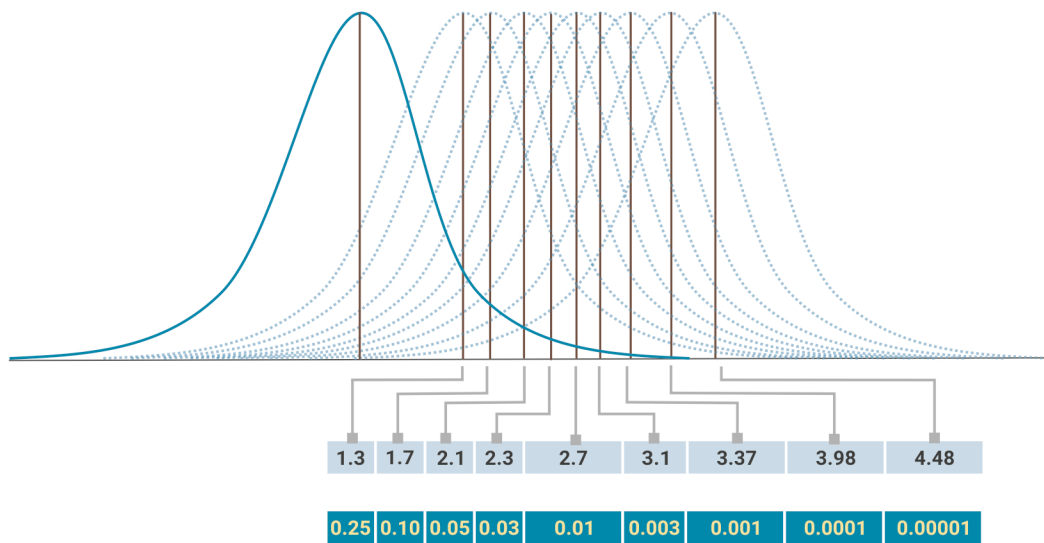


Figura 12.8.

Respuestas óptimas alcanzables bajo el supuesto de heredabilidad igual a 1 en función de diferentes fracciones de genotipos seleccionados. Los valores indicados en amarillo corresponden a la fracción seleccionada en la población (π), y los valores en negro a la ubicación de la media de la población (respuesta) expresado en desvíos estándares de la población original. Basado en Finney, 1958.

12.6.1.2. Selección en una etapa

Finney (1958) analizó las respuestas a la selección en un número variable de etapas. El autor definió primeramente a v como la relación entre los desvíos estándares fenotípico y genotípico de la distribución de N genotipos, o dicho de otra manera, el recíproco de la raíz cuadrada de la heredabilidad. Si N fuera suficientemente grande, la respuesta obtenible medida como diferencia entre el comportamiento (por ejemplo rendimiento de grano) de n genotipos seleccionados y los N iniciales podría ser mayor si los recursos

disponibles los aplicáramos a evaluar inicialmente no a la totalidad, sino solo a una fracción tomada al azar de N , designando a esa fracción como (α_0) . En consecuencia el número de genotipos para evaluar será $(\alpha_0 N)$.

Si el esquema de evaluación y consiguiente selección se concentra en una única etapa ($k=1$), entonces la proporción final de genotipos seleccionados será:

$$\alpha_0 \alpha_1 = \alpha$$

Finney (1958) ha probado una relación entre la respuesta obtenible en una etapa expresada en desvíos estándares y la proporción α_1 de genotipos seleccionados:

$$R_1 = (X_n - X_N) / \sigma_p \approx 1 / \alpha v^2 \approx h^2 / \alpha$$

donde:

R_1 es la respuesta a la selección en una única etapa,

X_n es la media de los genotipos seleccionados,

X_N es la media de todos los genotipos,

α es la fracción de genotipos seleccionados,

$v^2 = 1/h^2$, es la relación entre la variancia fenotípica y la variancia genotípica de la población de genotipos evaluados.

Las gráficas provistas por Finney (1958) permiten determinar para cada escenario definido por la magnitud (αv^2) cuál es la proporción α_1 que debería retenerse para acercarse a la máxima respuesta posible. Se recomienda al lector acudir al manuscrito original para inspeccionar la figura. Algunos conceptos que pueden extraerse de esta son los siguientes:

- Cuando la relación entre la variancia fenotípica y genotípica es elevada (alto v , baja h^2) la respuesta a la selección es reducida. En tal caso resultará conveniente hacer un descarte previo a la evaluación, de manera de evaluar $\alpha_0 N$ genotipos con un mayor nivel de recursos que aumenten la heredabilidad (que disminuyan v) al reducir el error.
- Cuando la relación entre la variancia fenotípica y genotípica es baja (bajo v , alta h^2), la respuesta esperada es más elevada y tanto mayor cuanto menor sea α_1 .

12.6.1.3. Selección en dos etapas

La selección en dos etapas ($k=2$) tiene ligeras ventajas respecto de la respuesta esperada con la selección en una única etapa cuando el error experimental es importante. El beneficio del descarte aleatorio anterior a la evaluación resulta de menor beneficio a medida que aumenta el número de etapas de selección. El óptimo de proporción de individuos retenidos en cada etapa es $\alpha_i = \alpha^{1/k}$, de modo que $\alpha_1 = \alpha_2 = \alpha_3 = \dots \alpha_k$.

Cochran (1951) analizó la selección en varias etapas empleando índices de selección y la manera de lograr una respuesta óptima. Si en la etapa 1 de la selección se retiene una proporción α_1 del total de genotipos, ello permitirá sumar parcelas de evaluación adicionales para cada genotipo en la etapa 2. Por ejemplo, si se dispusiera de 1000 parcelas para evaluar 500 genotipos y se seleccionara en la primera etapa $\alpha_1=0,50$ de los genotipos, para la etapa 2 se podrá disponer de $(1000/500)(1/\alpha_1) = 4$ repeticiones para evaluar, es decir, el doble que en la etapa 1.

El valor fenotípico del genotipo i -ésimo en la etapa 1 será:

$$P_{i1} = G_i + E_{i1} ,$$

donde G_i es el valor genotípico y E_{i1} es el desvío ambiental en la etapa 1 del genotipo i -ésimo.

La variancia fenotípica y la heredabilidad en la etapa 1 serán:

$$\sigma_{P1}^2 = \sigma_{G1}^2 + \sigma_{E1}^2$$

$$h^2 = \sigma_{G1}^2 / (\sigma_{G1}^2 + \sigma_{E1}^2)$$

Entonces: $\sigma_{P1}^2 = \sigma_{G1}^2 (1 + \sigma_{E1}^2 / \sigma_{G1}^2)$

Siguiendo a Baker (1986) designando a u como la relación $\sigma_{E1}^2 / \sigma_{G1}^2$, entonces:

$$\sigma_{P1}^2 = \sigma_{G1}^2 (1 + u)$$

En la segunda etapa, al haber más parcelas de evaluación, la variancia ambiental puede ser menor, en una proporción λ_1 que depende de la intensidad de selección utilizada en la primera etapa:

$$\sigma_{E2}^2 = \lambda_1 \sigma_{E1}^2 = \lambda_1 u \sigma_{G1}^2$$

y la variancia fenotípica en la etapa 2:

$$\sigma_{P2}^2 = \sigma_{G1}^2 (1 + \lambda_1 u)$$

Nótese que las variancias genéticas en etapa 1 y 2 son iguales, por cuanto estamos considerando que no hubo selección (solo un descarte al azar). En la etapa segunda, el fitomejorador puede utilizar un índice de selección cuyas ecuaciones serán:

$$b_1 \sigma_{G1}^2 (1 + \sigma_{E1}^2 / \sigma_{G1}^2) + b_2 \sigma_{G1}^2 = \sigma_{G1}^2$$

$$b_1 \sigma_{G1}^2 + b_2 \sigma_{G1}^2 (1 + \lambda_1 \sigma_{E1}^2 / \sigma_{G1}^2) = \sigma_{G1}^2$$

La resolución del sistema de ecuaciones es

$$b_1 = \lambda_1 u / (\lambda_1 u + u)$$

$$b_2 = u / (\lambda_1 u + u)$$

Si estos pesos del índice se expresan en forma relativa uno del otro, serán:

$$b_1 = \lambda_1$$

$$b_2 = 1$$

De modo que el índice fenotípico es:

$$I = \lambda_1 P_1 + P_2$$

Si se hubiese practicado selección en la etapa 1, entonces ocurrirá que $\sigma_{G1}^2 = \sigma_{G2}^2$, de manera que para estimar la respuesta a la selección basada en este índice es preciso tener en cuenta que una vez realizada la primera etapa de selección, habrá cambios en la variancia genotípica que queda a disposición para la segunda etapa (σ_{G2}^2) (Cochran, 1951). Consecuentemente, la variancia genética que se dispone para la segunda etapa de la selección es:

$$\sigma_{G2}^2 = \sigma_{G1}^2 [1 - h^2 i_1 (i_1 - t_1)]$$

donde

- σ_{G1}^2 es la variancia genotípica en la etapa 1,
- h^2 es la heredabilidad en la etapa 1,
- i_1 es la intensidad de selección en la etapa 1, y
- t_1 es el punto de truncamiento medido en desvíos estándar.

La variancia fenotípica en la segunda etapa será:

$$\sigma_{P2}^2 = \sigma_{G2}^2 + \sigma_{E2}^2$$

Cunningham (1975) demostró que la respuesta a la etapa 2 de la selección empleando el índice $I = \lambda_1 P_1 + P_2$ es:

$$R = i_1 \sigma_{G1} h + i_2 \sigma_{G1} \{(\lambda_1 + 1)(1 - h^2 s)\} / [(\lambda_1^2 (1 + u)(1 - s) + 2\lambda_1 (1 - s) + (1 - h^2 s + \lambda_1 u))^{0.5}],$$

donde

$$s = i_1 (i_1 - t_1)$$

De la aplicación de esta expresión de respuesta para diferentes situaciones de heredabilidad e intensidades de selección en la etapa 1 y 2 se puede concluir lo siguiente:

- Cualquiera será la heredabilidad, si la selección se realiza exclusivamente en una de las etapas, la respuesta esperada cuando se concentra solo en la etapa 2 es mayor que cuando se concentra solo en la etapa 1.
- Las mayores respuestas se logran cuando la proporción de individuos seleccionados se reparte entre ambas etapas, especialmente cuando la intensidad de selección es mayor en la etapa primera (quedarán más parcelas de evaluación por genotipo en la etapa 2.) La ventaja de la selección en dos etapas respecto a una sola es mayor cuando las heredabilidades son bajas a intermedias ya que el descarte en la etapa 1 por selección libera más parcelas de evaluación para la segunda etapa.
- El efecto favorable de dividir la proporción de individuos seleccionados entre las dos etapas parece ser mayor cuanto menor es la heredabilidad en la etapa 1.
- En todos los casos, como era de prever, las mayores respuestas se logran con heredabilidades mayores.

La respuesta total a la selección suele ser mayor cuando $\alpha_1 = \alpha_2 = \alpha^{0.5}$ para varios valores de h^2 y donde α es la proporción total de genotipos seleccionados. De acuerdo a Cunningham (1975), si se hicieran 3 etapas, la mejor opción sería $\alpha_1 = \alpha^{1/3}$

Alternativamente, podría haberse planteado seleccionar en la etapa 1 y para la selección en la etapa 2 tener en cuenta en forma conjunta las evaluaciones de ambas etapas. Para ello puede definirse un índice de selección de la siguiente forma:

$$I = b_1 P_1 + b_2 P_2 ,$$

donde

P_1 y P_2 son las medias fenotípicas de los genotipos en las etapas 1 y 2.

Los coeficientes b_i son los coeficientes de regresión parcial de las medias fenotípicas en las etapas 1 y 2 con el valor genotípico g , respectivamente.

Lo que importa es la relación b_1/b_2 , a los que se los puede estandarizar de modo que su sumatoria sea 1. Puede demostrarse que los pesos del índice óptimo dependen del número de repeticiones (Wricke y Weber, 1986)

$$b_1 = r_1 / (r_1 + r_2),$$

$$b_2 = r_2 / (r_1 + r_2),$$

donde r_1 y r_2 son el número de repeticiones en cada etapa.

12.7. Selección con base en información de parientes

Consideremos una población estructurada en f familias, cada una de ellas integrada por igual número de individuos.

Por una parte el valor fenotípico promedio de la f -ésima familia obedece a la suma de su valor genotípico (G_f) y el desvío ambiental (error) correspondiente (E_f):

$$P_f = G_f + E_f$$

Por otra parte dentro de cada familia, para cada k -ésimo individuo que la compone es posible calcular un desvío entre el valor fenotípico promedio de la familia y el valor fenotípico del k -ésimo individuo (P_d):

$$P_f - P_d,$$

donde

$$P_d = G_d + E_d$$

Siendo G_d el valor genotípico del k -ésimo individuo y E_d el desvío ambiental correspondiente.

El objetivo es mejorar el genotipo agregado definido por la suma de los valores genotípicos promedio de la familia (G_f) y del k -ésimo individuo (G_d).

$$H = G_f + G_d,$$

empleando un índice fenotípico de la forma (Baker, 1986):

$$I = b_f P_f + b_d P_d$$

Las ecuaciones del índice son:

$$b_f \sigma_{P_f}^2 + b_d \sigma_{P(f,d)} = \sigma_{G_f}^2 + \sigma_{G(f,d)}$$

$$b_f \sigma_{P(f,d)} + b_d \sigma_{P_d}^2 = \sigma_{G(f,d)} + \sigma_{G_d}^2$$

Como los términos de covarianza son iguales a cero, el sistema y su solución serán:

$$b_f \sigma_{P_f}^2 = \sigma_{G_f}^2$$

$$b_d \sigma_{P_d}^2 = \sigma_{G_d}^2$$

$$b_f = \sigma_{G_f}^2 / \sigma_{P_f}^2 = h_f^2$$

$$b_d = \sigma_{G_d}^2 / \sigma_{P_d}^2 = h_d^2$$

Denominando h^2 a la heredabilidad sobre la base de plantas individuales, las heredabilidades entre familias (h_f^2) y dentro de familias (h_d^2) serán, respectivamente:

$$h_f^2 = h^2 [1 + (n - 1)r] / [1 + (n - 1)t]$$

$$h_d^2 = h^2 (1 - r) / (1 - t),$$

donde:

r es el coeficiente de correlación intraclase genotípico para el tipo de familia utilizado y

t es el coeficiente de correlación intraclase fenotípico para el tipo de familia utilizado.

Para el caso de familias de hermanos completos, $r_{HC} = 1/2$ y para familias de medios hermanos, $r_{MH} = 1/4$.

El coeficiente de correlación intraclase fenotípico para el tipo de familia utilizado,

$$t = (r\sigma_G^2 + \sigma_E^2/n) / (\sigma_G^2 + \sigma_E^2)$$

Cuando n es suficientemente grande, t tiende a ser igual a rh^2 .

Finalmente, $I = h_f^2 P_f + h_d^2 P_d$.

La respuesta a la selección combinada se expresa como

$$R_c = i\sigma_G h [1 + r^2(i - h^2)(n - 1) / 1 + (n - 1)rh^2]^{0.5}.$$

12.8. Métodos para la elección de progenitores basados en caracteres múltiples

Considerando que la mayoría de los caracteres cuantitativos son controlados por muchos loci de pequeño efecto y con acción aditiva, es posible estimar o predecir el comportamiento promedio poblacional a partir de las características genéticas de los posibles genotipos progenitores de esa población. El análisis vectorial y el análisis canónico pueden utilizarse para elegir los progenitores de una población de cría de modo tal que presente el mayor comportamiento medio posible entre diferentes alternativas. Por el contrario, mediante métodos basados en *distancias multivariadas* pueden elegirse aquellos progenitores que permitan maximizar la variabilidad genética de la población de cría que se desea generar. Cuando los objetivos del programa de mejoramiento son a corto plazo, el análisis canónico y el análisis vectorial serían los más apropiados, mientras que en programas de largo plazo donde interesa disponer de un elevado nivel de variabilidad genética, los métodos basados en distancias generalizadas serían los más convenientes (Baker, 1986). Cualquiera sea el método adoptado, si la epistasis fuera importante, la eficiencia de estos sería menor que en su ausencia.

El análisis vectorial fue propuesto por Grafius (1965) y requiere considerar qué caracteres tener en cuenta y cuáles son sus importancias relativas. A diferencia de los pesos económicos de los índices de selección, Grafius propone utilizar los coeficientes de regresión múltiple estandarizados correspondientes a distintos caracteres, considerados como variables independientes, en relación con una estimación del mérito agregado (variable dependiente). Los valores fenotípicos de los m caracteres deben transformarse de manera de tener una media general igual a 1. En cierto modo, el método implica la definición de un ideotipo, donde los caracteres considerados se ponderan por su importancia relativa. A cada potencial progenitor corresponderá un vector y también habrá un vector que define el ideotipo o genotipo ideal (Figura 12.9). El método de Grafius no solo permite establecer cuáles deben ser los progenitores de la población de cría, sino también en qué proporción deberían contribuir a esta.

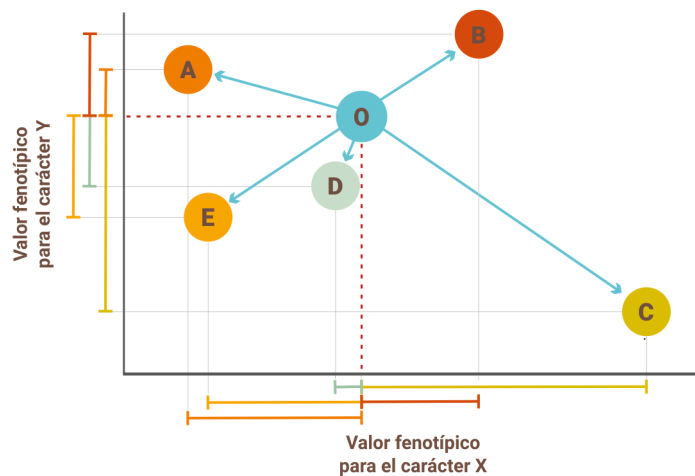
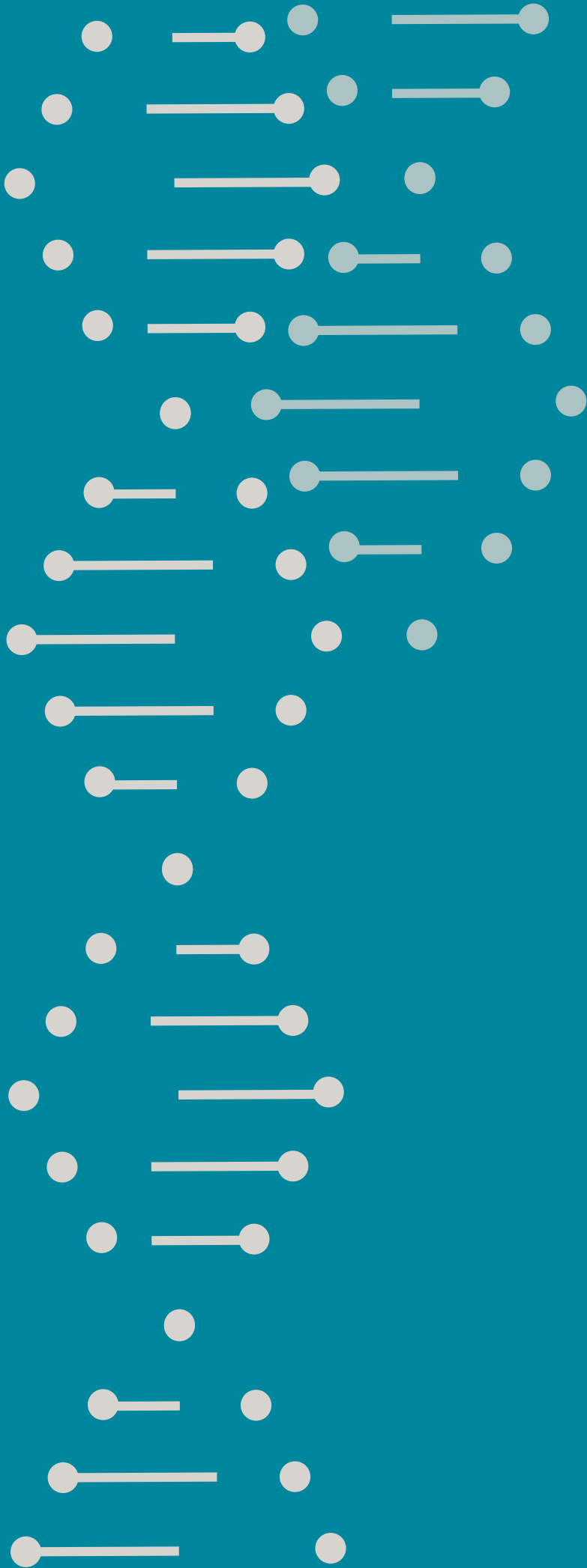


Figura 12.9. Representación de la ubicación de cinco eventuales progenitores (A, B, C, D y E) de una población de cría y de un genotipo definido como óptimo (O) en un plano determinado por los valores fenotípicos para dos caracteres de interés, X e Y.

Una variante del método vectorial es el análisis propuesto por Pederson (1981).

El análisis canónico es un método multivariado que representa los datos fenotípicos de los potenciales progenitores de la población de cría de modo tal que permita la mayor discriminación entre ellos o sus agrupamientos. La primera variable canónica se calcula transformando los datos para maximizar la relación entre la variancia entre grupos respecto a la variancia dentro de grupos. El segundo y posteriores ejes canónicos siguen en importancia decreciente a esa relación. Al igual que en los métodos anteriores, los progenitores se eligen teniendo en cuenta la media de la población que podría formarse y su desvío con la de la población ideal. En este método no hay asignación de pesos o importancia relativa a los caracteres y solo se tienen en cuenta por su poder discriminante entre los padres.

Finalmente los métodos basados en distancias generalizadas estiman las que existen entre un grupo de potenciales progenitores de una población para varios caracteres de interés. Es un método apto para elegir progenitores de poblaciones de cría de especies autóгамas.





Interacción genotipo x ambiente

13.1. Introducción

En el Capítulo 3 presentamos un modelo lineal sencillo que expresa el valor fenotípico (P) como la suma de dos términos, el valor genotípico (G) y el desvío ambiental (E):

$$P = G + E, \quad [3.1]$$

La sola sumatoria del valor genotípico y el desvío ambiental puede no explicar suficientemente el valor fenotípico de un número de genotipos evaluados en diferentes ambientes. En tales situaciones debe incluirse otro término que de cuenta del comportamiento diferencial de los genotipos cuando se los expone a una serie de ambientes. Como veremos más adelante, y dependiendo del modelo, ese término puede analizarse conside-

rando una relación de efectos multiplicativos entre genotipos y ambiente. En tal caso puede encontrarse una analogía con lo que sucede entre un carácter complejo y sus componentes (Baker, 1986). Afortunadamente los apartamientos de la linealidad no son tan drásticos como para afectar la eficiencia de modelos que parten de ese supuesto.

13.2. Modelo con interacción genotipo x ambiente

El modelo [3.1] asume que todos los genotipos responden de manera similar a modificaciones en el ambiente en el que se desenvuelven, supuesto que frecuentemente no se cumple. Cuando los genotipos se comportan de modo diferenciado a las condiciones del ambiente, los valores fenotípicos responden a los componentes causales anteriores, más el efecto de la interacción genotipo x ambiente. De tal manera:

$$P = G + E + IGA$$

El término de interacción, IGA, no debe confundirse con la covariancia entre los valores genotípicos y los desvíos ambientales. Esta última es producto de la falta de independencia entre ambos términos y puede ser minimizada por el experimentador mediante la simple aleatorización de los genotipos a las unidades experimentales disponibles. Por el contrario, el efecto de la interacción genotipo x ambiente es una propiedad inherente al genotipo, que contribuye a “obscurecer” aún más la correspondencia entre valor fenotípico y valor genotípico para los caracteres cuantitativos. Ello significa que el fitomejorador debe tener en cuenta los efectos de interacción genotipo x ambiente, especialmente en programas de selección.

En el contexto de la tarea del mejoramiento genético al referirnos al ambiente (o macroambiente) aludimos al efecto de localidad, año, estación de crecimiento, manejo del cultivo, condiciones de una prueba a escala de laboratorio, etc. que imperan sobre el conjunto de genotipos que se están evaluando o comparando. Puede definirse como microambiente a las condiciones en las que un individuo o genotipo en particular crece rodeado de otros en una unidad experimental. Las condiciones microambientales son mucho más difíciles de controlar, aun a escala de laboratorio.

A fin de clarificar el concepto de la interacción genotipo x ambiente, la ejemplificaremos en el caso más sencillo de comparar el comportamiento de solo dos genotipos (1 y 2) expuestos a dos ambientes (A y B) (Tabla 13.1).

Genotipos	Ambientes		Diferencia	Promedio a través de ambientes
	A	B	A-B	
1	P_{1A}	P_{1B}	$P_{1A} - P_{1B}$	$(P_{1A} + P_{1B})/2 = P_{1.}$
2	P_{2A}	P_{2B}	$P_{2A} - P_{2B}$	$(P_{2A} + P_{2B})/2 = P_{2.}$
Diferencia 1-2	$P_{1A} - P_{2A}$	$P_{1B} - P_{2B}$	$P_{1A} - P_{1B} - P_{2A} + P_{2B}$	
Promedio a través de genotipos	$(P_{1A} + P_{2A})/2 = P_{.A}$	$(P_{1B} + P_{2B})/2 = P_{.B}$		$(P_{1A} + P_{2A} + P_{1B} + P_{2B})/4 = P_{..}$

Tabla 13.1.

Análisis del comportamiento de dos genotipos (1 y 2) expuestos a dos condiciones ambientales (A y B). P_{ij} son los valores fenotípicos observados para el genotipo i-ésimo en el ambiente j-ésimo.

La existencia de interacción genotipo x ambiente puede comprobarse observando si las diferencias de comportamiento de los dos genotipos expuestos a dos ambientes son distintas de cero. En la Tabla 13.1 ese valor se presenta en la celda con texto destacado en negrita. Si dicha celda es distinta de cero, ello sucede cuando:

$$(P_{1A} - P_{1B}) \neq (P_{2A} - P_{2B}), \text{ o su equivalente:}$$

$$(P_{1A} - P_{2A}) \neq (P_{1B} - P_{2B})$$

Los términos de cada lado de la desigualdad pueden tener ambos el mismo signo (>0 o <0) o diferente signo. Esto permite diferenciar dos tipos de interacciones genotipo x ambiente. En un primer caso se trata de interacciones cuantitativas o no cruzadas (o de tipo I) y en el otro caso cualitativas o cruzadas (o de tipo II). La denominación de una u otra hace referencia a su representación en un gráfico de coordenadas, donde en el eje de las abscisas se representan a los valores ambientales y en el eje de las ordenadas a los valores fenotípicos (Figura 13.1). Existen pruebas estadísticas para determinar si las interacciones genotipo x ambiente entre pares de genotipos evaluados en pares de ambientes son de tipo cruzado (Azzalini y Cox, 1984).

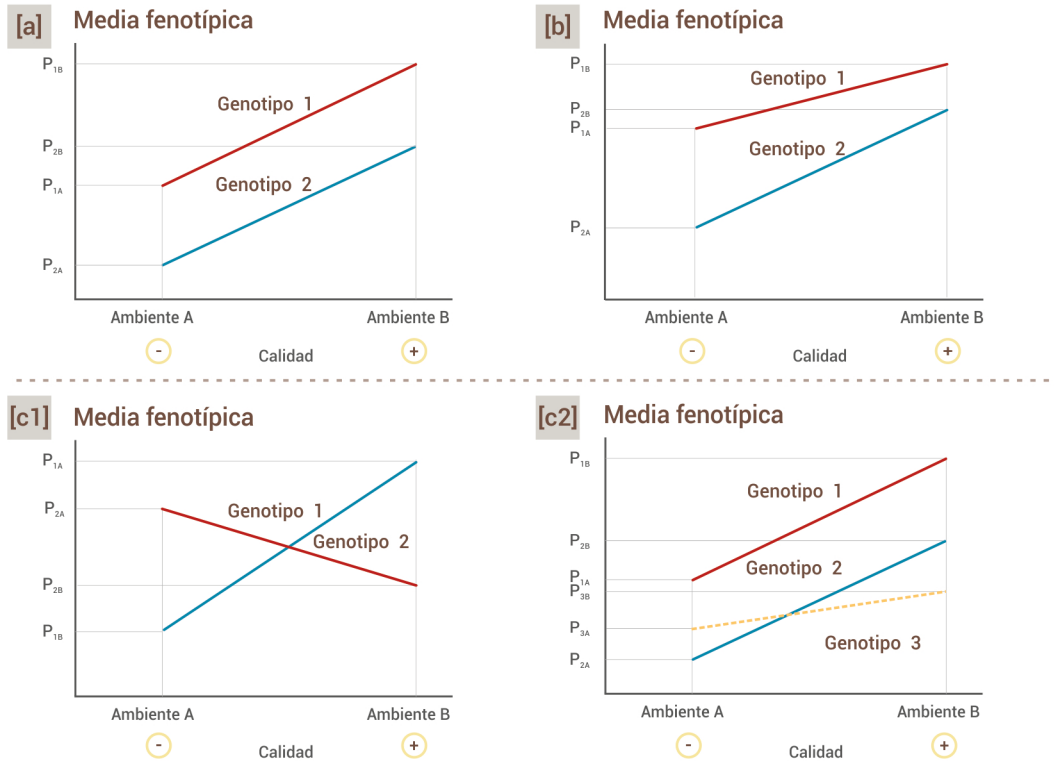


Figura 13.1.

Tipos de interacciones genotipo x ambiente. Entre dos genotipos: [a] Ausencia de interacción; [b] Interacción cuantitativa o no cruzada; [c1] Interacción cualitativa o cruzada. Entre tres genotipos: [c2] Interacción cruzada entre Genotipos 2 y 3, interacción no cruzada entre Genotipos 1 y 3, ausencia de interacción entre Genotipos 1 y 2.

Para el caso más simple de interacción entre dos genotipos evaluados en los dos ambientes, aplicaremos el siguiente modelo:

$$P_{ij} = M + G_i + E_j + IGA_{ij}$$

donde:

M es la media general,

P_{ij} es la media del valor fenotípico del i -ésimo genotipo en el j -ésimo ambiente,

G_i es el efecto del i -ésimo genotipo,

E_j es el efecto del j -ésimo ambiente,

IGA_{ij} es el efecto de interacción genotipo x ambiente del i -ésimo genotipo evaluado en el j -ésimo ambiente.

A manera de ejemplo, si se evalúan los genotipos 1 y 2 en los ambientes A y B, y las medias observadas a través de repeticiones fueran las indicadas en la Tabla 13.2, es posible calcular los efectos considerados en el modelo.

Genotipos	Ambientes		Diferencia	Promedio
	A	B	A-B	
1	100	140	-40	120
2	80	100	-20	90
Diferencia 1-2	20	40	-20	
Promedio	90	120		105

Tabla 13.2. Análisis del comportamiento de dos genotipos (1 y 2) expuestos a dos condiciones ambientales (A y B). Los números descaltados en negrita cursiva corresponden a valores fenotípicos observados promedio para el genotipo i -ésimo en el ambiente j -ésimo (P_{ij}).

Los efectos de cada ambiente medidos como desvío de la media general son los siguientes:

$$E_A = 90 - 105 = -15$$

$$E_B = 120 - 105 = 15$$

Análogamente, los efectos de cada genotipo medidos como desvío de la media general serán:

$$G_1 = 120 - 105 = 15$$

$$G_2 = 90 - 105 = -15$$

Finalmente los efectos de interacción genotipo x ambiente pueden calcularse por diferencia de acuerdo al siguiente modelo:

$$IGA_{ij} = P_{ij} - (M + G_i + E_j)$$

De tal manera:

$$IGA_{1A} = 100 - (105 + 15 - 15) = -5$$

$$IGA_{2A} = 80 - (105 - 15 - 15) = 5$$

$$IGA_{1B} = 140 - (105 + 15 + 15) = 5$$

$$IGA_{2B} = 100 - (105 - 15 + 15) = -5$$

Por lo tanto los valores fenotípicos del genotipo 1 en los ambientes A y B pueden expresarse como:

$$P_{1A} = 100 = 105 + 15 - 15 - 5$$

$$P_{1B} = 140 = 105 + 15 + 15 + 5$$

En el ejemplo que sigue, las interacciones son de tipo cruzado (Tabla 13.3):

Genotipos	Ambientes		Diferencia	Promedio
	A	B	A-B	
1	100	80	20	90
2	80	160	-80	120
Diferencia 1-2	20	-80	100	
Promedio	90	120		105

Tabla 13.3.

Análisis del comportamiento de dos genotipos (1 y 2) expuestos a dos condiciones ambientales (A y B). Los números descaltados en negrita cursiva corresponden a valores fenotípicos observados promedio para el genotipo i -ésimo en el ambiente j -ésimo (P_{ij}).

Los efectos de cada ambiente medidos como desvío de la media general son los siguientes:

$$E_A = 90 - 105 = -15$$

$$E_B = 120 - 105 = 15$$

Los efectos de los genotipos medidos como desvío de la media general son análogamente:

$$G_1 = 90 - 105 = -15$$

$$G_2 = 120 - 105 = 15$$

Los efectos de interacción genotipo x ambiente pueden calcularse por diferencia de acuerdo al siguiente modelo:

$$IGA_{ij} = P_{ij} - (M + G_i + E_j)$$

De tal manera:

$$IGA_{1A} = 100 - (105 - 15 - 15) = 25$$

$$IGA_{2A} = 80 - (105 + 15 - 15) = -25$$

$$IGA_{1B} = 80 - (105 - 15 + 15) = -25$$

$$IGA_{2B} = 160 - (105 + 15 + 15) = 25$$

A partir de los efectos puede reproducirse el valor fenotípico del genotipo 1 en los ambientes A y B de acuerdo al modelo genético:

$$P_{1A} = 100 = 105 - 15 - 15 + 25$$

$$P_{1B} = 80 = 105 - 15 + 15 - 25$$

Desde el punto de vista del fitomejorador, no solo importa conocer la existencia de efectos de interacción genotipo x ambiente, sino también si son de tipo cruzado o no cruzado, y cuáles son los factores del ambiente involucrados en el efecto de interacción, especialmente si son controlables o no. En el caso de que los factores ambientales que generan la interacción sean controlables por el agricultor, como por ejemplo el manejo del cultivo (nutrición, fecha de siembra, control de una enfermedad), el fitomejorador puede tenerlo en cuenta al momento de seleccionar específicamente para esa condición de cultivo. En este caso el fitomejorador trataría de generar genotipos específicamente adaptados a condiciones ambientales controlables. Cuando los factores ambientales que provocan la interacción resultan incontrolables, el fitomejorador orienta su actividad buscando genotipos que posean un comportamiento más estable o predecible en un rango ambiental compatible con el que se supone estará expuesto el cultivar una vez liberado al mercado.

El objetivo de los programas de mejoramiento es obtener el mejor genotipo para un determinado rango ambiental o megambiental. El comportamiento diferencial de una serie de genotipos puede servir como criterio para determinar si en un amplio rango de ambientes es posible delimitar subambientes. Tal delimitación debería establecerse de manera tal que dentro de cada subambiente solo se espere que existan interacciones cuantitativas, correspondiendo a cada uno de ellos un genotipo ganador. Si por el contrario se detectaran interacciones cruzadas o cualitativas, ello sería evidencia de la existencia de subambientes, cada uno de ellos con un genotipo ganador distinto. En tales situaciones merecería analizarse la justificación de contemplar el desarrollo de cultivares para distintos subambientes.

Obsérvese que en cualquier caso las medidas de interacción de un genotipo se definen en un marco conceptual dado por el resto de los genotipos que se evalúan junto con él y el grupo de ambientes analizado, es decir, que las propiedades o características que se atribuyan a un genotipo en cuanto a sus efectos de interacción con el ambiente pueden variar si se hubieran considerado otros genotipos de cotejo u otros ambientes.

Cuando los genotipos de interés se evalúan en un único ambiente, no es posible determinar la existencia de variancia para los efectos de interacción genotipo x ambiente. La única manera de poder determinarlo es evaluar a los genotipos en un número de ambientes igual o mayor que dos. Cuanto mayor sea el número de ambientes, menos sesgadas resultarán las estimaciones de las variancias genéticas. Por lo tanto debe tenerse en cuenta que las estimaciones de variancia genética y de heredabilidad obtenidas de un único experimento resultarán en sobreestimaciones de los verdaderos parámetros poblacionales.

13.3. Interacción genotipo x ambiente y predicción de la respuesta a la selección

La ecuación de predicción de respuesta a la selección presentada en los Capítulos 9 y 10 asume que los componentes de variancia que se utilizan están estimados sin sesgo. Sin embargo para que ello ocurra debemos haber dispuesto de una población (o muestra) de ambientes y de genotipos suficientemente grandes. Esto es infrecuente, especialmente en el caso de ambientes. La consecuencia de ello es que la covariancia entre valores genotípicos y desvíos ambientales (ver Capítulo 3) puede ser diferente de cero.

En estos casos, Helms y Hammond (2006) presentaron modificaciones a las expresiones clásicas de la heredabilidad (h^2) y de la respuesta a la selección (R) en e ambientes y b repeticiones por ambiente:

$$h^2 = \sigma_A^2 / [\sigma_G^2 + (\sigma_{GE}^2 + 2\sigma_{GE})/e + \sigma^2/be]$$

donde:

σ_A^2 es la variancia aditiva de las unidades de selección,

σ_G^2 es la variancia genotípica de las unidades de selección,

σ_{GE}^2 es la variancia de interacción entre genotipos y ambientes,

σ_{GE} es la covariancia entre genotipos y ambientes,

σ^2 es la variancia del error experimental.

En la medida que la covariancia entre genotipos y ambientes sea diferente de cero, la heredabilidad será menor a la que se esperaría empleando un número grande de ambientes.

La respuesta esperada a la selección será:

$$R = i r \sigma_A \{ \sigma_A^2 / [\sigma_G^2 + (\sigma_{GE}^2 + 2\sigma_{GE})/e + \sigma^2/be] \}^{0.5}$$



donde:

i es la intensidad de selección,

r es la correlación entre los ambientes de evaluación y la población de ambientes a la que se destinarán los genotipos seleccionados,

σ_A es el desvío estándar de la variancia genética en la población ambiental a la que estarán destinados los genotipos seleccionados.

Desde un punto de vista práctico, el efecto del sesgo en las estimaciones de los componentes de variancia no sería muy significativo cuando los genotipos se seleccionan con base en evaluaciones en varias localidades y años.

El efecto más importante sobre la convergencia de las respuestas real y predicha viene dado por la correlación (r) que exista entre los ambientes de prueba y la población ambiental a la que se destinarán los genotipos (Helms y Hammond, 2006).

13.4. Experimentos multiambientales

Las fuentes de variación en experimentos replicados en una serie de ambientes pueden ser distintas según cuál sea la necesidad de las evaluaciones y los recursos disponibles. Cada situación se describe con su modelo estadístico específico (Tabla 13.4).

Evaluación en...	Modelo
Única localidad o ambiente	$P_{ij} = u + b_j + g_i + e_{ij}$
Varias localidades o ambientes	$P_{ijk} = u + a_j + b_{k(j)} + g_i + ga_{ij} + e_{ijk}$
Varios años en un ambiente	$P_{ijk} = u + y_m + b_{k(m)} + g_i + gy_{im} + e_{imk}$
Varios años y localidades	$P_{ijkm} = u + y_m + a_j + ya_{mj} + b_{k(mj)} + g_i + gy_{im} + ga_{ij} + gya_{imj} + e_{ijmk}$

Tabla 13.4.

Modelo estadístico correspondiente a diferentes estrategias de evaluación en un ambiente único y en evaluaciones multiambientales. b_j es el efecto de la j -ésima repetición en un experimento llevado a cabo en un único ambiente; g_i es el efecto del i -ésimo genotipo; e_{ij} es el error experimental correspondiente a la unidad experimental con el i -ésimo genotipo en la j -ésima repetición; e_{ijk} es el error experimental asociado a la unidad experimental con el i -ésimo genotipo, en el j -ésimo ambiente o localidad y k -ésima repetición dentro de ese ambiente; e_{imk} es el error experimental asociado a la unidad experimental con el i -ésimo genotipo, en el m -ésimo año de evaluación y la k -ésima repetición dentro de ese año; e_{ijmk} es el error experimental asociado a la unidad experimental con el i -ésimo genotipo, en el ambiente j -ésimo dentro del m -ésimo año de evaluación y la k -ésima repetición dentro de ese año y localidad.

Las consideraciones referidas a efectos fijos y efectos aleatorios realizadas en capítulos previos y en el Anexo son de aplicación en el caso de evaluaciones multiambientales. Cuando las evaluaciones se realizan en un número limitado de años y localidades, puede ser más conveniente agrupar ambas fuentes de variación directamente como “ambientes”, sin descomponerlas en año, localidad e interacción año x localidad y considerar a los ambientes de prueba como una muestra de un factor aleatorio. Conceptualmente esto significa que los años y localidades de evaluación utilizados constituyen una muestra aleatoria de una población ambiental que posee cierta variabilidad espacial y temporal similar a la que podrán verse expuestos los genotipos una vez liberados al gran cultivo. Más recientemente cuando se emplean modelos mixtos para estimar variancias, si el número de ambientes de prueba es reducido, puede ser más conveniente considerar el factor ambiente como fijo para mejorar las estimaciones de las variancias de genotipos y de interacción genotipo x ambiente.

La experiencia indica que las interacciones genotipo x año suelen ser relativamente más importantes que las interacciones genotipo x localidad, pero ello depende de las particularidades del rango de ambientes que se está muestreando. En zonas llanas, en general, puede ocurrir que las interacciones genotipo x localidad sean menos importantes que las interacciones genotipo x año, especialmente cuando existe variabilidad climática interanual que condiciona la disponibilidad hídrica en etapas críticas del cultivo y la ocurrencia de adversidades bióticas. Por el contrario, en zonas de relieve, diferencias de altura sobre el nivel del mar, pendientes y orientación pueden ser importantes en localidades distantes pocos kilómetros, por lo que la interacción genotipo x localidad puede tener una incidencia similar a la interacción genotipo x año. La naturaleza de la adversidad o de factor que limita el rendimiento también influye sobre el tipo de interacción genotipo x ambiente. Por ejemplo, las interacciones de tipo cruzado son frecuentemente significativas cuando se evalúan genotipos con distinto nivel de tolerancia a enfermedades y plagas en ambientes donde el factor biótico que resulta limitante está presente o ausente.

Cuando existe interacción genotipo x ambiente, la comparación de genotipos depende de la definición de un dominio ambiental, el que a su vez delimita el alcance de la recomendación y selección de genotipos, los estudios de herencia, de heredabilidad y la identificación de QTLs. En los programas de mejoramiento genético es importante conocer los efectos de interacción genotipo x ambiente para delimitar megambientes, interpretar correctamente los resultados de evaluaciones y establecer la amplitud y estructura necesarias de la etapa de evaluación de genotipos o *testing*. La delimitación de los megambientes se realiza minimizando las interacciones genotipo x ambiente dentro de ellos y maximizándolas entre ellos. Cuanto más importante sea la interacción genotipo por ambiente, mayor será la exigencia o magni-

tud del programa de *testing* o evaluación. La delimitación de megambientes permite la organización de un sistema de evaluación de manera de distribuir los ambientes de prueba en forma más eficiente, evitando evaluar en ambientes similares que proveen información redundante, y en cambio hacerlo en condiciones que permitan la manifestación de las interacciones genotipo x ambiente y selección por comportamiento promedio y estabilidad. Alternativamente, dependiendo de la extensión de un megambiente, este puede constituirse como el dominio ambiental para el que se desarrollarán cultivos, y la magnitud del sistema de testing puede ser reducido, ya que bajo los criterios enunciados anteriormente, dentro de cada megambiente no se esperarán interacciones genotipo x ambiente significativas.

Frecuentemente ocurre que en una serie de experimentos multiambientales, el efecto principal del ambiente puede ser muy importante como para confundir u oscurecer el análisis del patrón de respuesta de los genotipos evaluados frente a variaciones importantes en la calidad ambiental. Para la selección de genotipos en estas situaciones Fox y Rosielle (1982) demostraron la conveniencia de estandarizar los datos originales previo al análisis de agrupamiento de ambientes en función de sus características.

La distancia euclidiana ($d_{jj'}^2$) (ver Anexo) entre dos ambientes cualesquiera, j y j' , medida por el comportamiento de g genotipos será:

$$d_{jj'}^2 = \sum_{i=1}^g (y_{ij} - y_{ij'})^2 = \sum_{i=1}^g (g_i + a_j + ga_{ij} - g_i - a_{j'} - ga_{ij'})^2 = \sum_{i=1}^g [(a_j - a_{j'}) + (ga_{ij} - ga_{ij'})]^2$$

donde:

y_{ij} , $y_{ij'}$ son los valores fenotípicos del genotipo i -ésimo en los ambientes j -ésimo y j' -ésimo,

g_i es el valor genotípico del i -ésimo genotipo,

a_j , $a_{j'}$ son los efectos del ambiente j -ésimo y j' -ésimo,

ga_{ij} , $ga_{ij'}$ son los efectos de interacción entre el genotipo i -ésimo y el ambiente j -ésimo y j' -ésimo, respectivamente.

Cuando las diferencias entre las medias ambientales son muy grandes en relación con las diferencias debidas a las interacciones, el efecto principal de ambientes será proporcionalmente mayor que el efecto de las interacciones. La estandarización de los datos fenotípicos permite reducir la distorsión resultante de la mayor importancia del efecto principal de ambientes. La estandarización de los valores fenotípicos observados para cada genotipo en cada ambiente, previo al cálculo de las distancias, hace que las medias ambientales sean iguales a cero y los desvíos estándar iguales a uno para todos ellos.

Las medias genotípicas estandarizadas del *i*-ésimo genotipo se obtienen sustrayendo de la media observada en cada ambiente (g_{ij}) su media genotípica a través de ambientes (g_i) y dividiéndola por su desvío estándar (S_i). La correlación entre las medias estandarizadas de todos los genotipos en cada par de ambientes (r_{ij}) es una medida del parecido entre ese par de ambientes. La distancia euclidiana entre dos ambientes medida utilizando estandarización de los valores fenotípicos observados equivale a la expresión ' $d_{ij}^2 = 2(g-1)(1-r_{ij})$ ', donde g es el número de genotipos evaluados (Fox y Rosielle, 1982).

13.5. Modelos estadísticos de análisis de la interacción genotipo x ambiente

Cuando se realizan experimentos multiambientales, el análisis de la interacción genotipo x ambiente comienza con un análisis de variancia (ANDeVA) por ambiente y a través de ambientes o combinado. Asumamos que se evalúan en experimentos de bloques completos aleatorizados con tres repeticiones a 20 genotipos en 10 ambientes. Los ANDeVA por ambiente y combinado o a través de ambientes tendrán las configuraciones que se presentan en las Tabla 13.5 y 13.6.

Fuentes de variación	Grados de libertad	Cuadrado medio	Fc
Repeticiones	$r-1 = 2$	CM_r	CM_r/CM_{ee}
Genotipos	$g-1 = 19$	CM_g	CM_g/CM_{ee}
Error experimental	$(r-1)(g-1) = 38$	CM_{ee}	
Total	$rg-1 = 59$		

Tabla 13.5. Esquema del análisis de variancia de 20 genotipos en cada ambiente siguiendo un diseño de bloques completos aleatorizados con tres repeticiones.

Fuentes de variación	Grados de libertad	Cuadrados medios	F _c
Ambientes	$a-1 = 9$		
Repeticiones (ambientes)	$a(r-1) = 20$		
Genotipos	$g-1 = 19$	CM_g	CM_g/CM_{gxa}
Genotipo x Ambientes	$(g-1)(a-1) = 171$	CM_{gxa}	CM_{gxa}/CM_{eec}
Error experimental	$a(r-1)(g-1) = 380$	CM_{eec}	
Total	$arg-1 = 599$		

Tabla 13.6. Esquema del análisis de variancia de 20 genotipos evaluados en 10 ambientes siguiendo un diseño de bloques completos aleatorizados con tres repeticiones.

Puede ocurrir que las variancias de los errores entre experimentos sean heterogéneas (diferentes). Para probarlo se emplea el test de Bartlett (1937) (ver Anexo). Un indicio rápido de ocurrencia de heterogeneidad de la variancia del error es que el cociente entre el mayor y el menor cuadrado medio del error experimental encontrados en los análisis de variancia por ambiente sea mayor o igual a tres. Las pruebas de significancia de genotipos y de la interacción genotipo x ambiente deben considerar la homogeneidad o heterogeneidad de los errores experimentales, como se indica en la Tabla 13.7.

		Denominador de pruebas de significancia de fuentes de variación	
CM _{ee}	Interacción Genotipo x Ambiente (IGA)	Genotipos	
		IGA ausente	IGA presente
		CM _{eecc'} , a(r-1)(g-1) grados de libertad	
Homogéneos	CM _{eecc'} , a(r-1)(g-1)	[(SC _{gxa} + SC _{eecc'})/(gl _{ga} + gl _{eecc'}), (gl _{ga} + gl _{eecc'}) grados de libertad	CM _{gxa'} , (a-1)(g-1)
Heterogéneos	CM _{eecc'} , grados de libertad del exp. con mayor CM _{ee'} , (r-1)(g-1)	[(SC _{gxa} + SC _{eecc'})/(gl _{ga} + gl _{eecc'}), grados de libertad del exp. con mayor CM _{ee'} , (r-1)(g-1)	CM _{gxa} , (a-1)(g-1)

Tabla 13.7.

Denominador y grados de libertad a utilizar en las pruebas de significancia de los cuadrados medios de interacción genotipo x ambiente (CM_{eecc'}) y de genotipos (CM) en análisis multiambiental de g genotipos evaluados en a ambientes con r repeticiones. CM_{ee'} y CM_{eecc'} representan los cuadrados medios del error experimental de experimentos individuales y del análisis combinado a través de a ambientes. Se asume que todos los experimentos tienen las mismas dimensiones y que las interacciones genotipo x ambiente son homogéneas.

Según ello, asumiendo que las variancias o cuadrados medios de los errores entre experimentos fueran homogéneas, el denominador de la prueba de significancia de la interacción genotipo x ambiente debería ser el cuadro medio del error experimental combinado. Si por el contrario, las variancias de los errores experimentales fueran heterogéneas, debería emplearse una prueba de significancia de la interacción genotipo x ambiente más conservadora. Para ello se utilizará el mismo cuadrado medio del error experimental combinado, pero contrastando con el valor de F tabulado con los grados de libertad de la interacción genotipo x ambiente, y los grados de libertad del error experimental correspondiente al experimento que mostró mayor variancia del error.

La modalidad de análisis de las interacciones genotipo x ambiente puede ser univariada o multivariada. En el primer caso, la respuesta fenotípica de los genotipos en diferentes ambientes se considera como la expresión de

un mismo carácter en función de diferentes calidades de ambiente. Se trata de un modelo de regresión lineal simple, donde para cada genotipo se determinan la pendiente de la recta de respuesta y los desvíos de la regresión (Eberhart y Russell, 1966).

En el modelo de Eberhart y Russell (1966), la partición de la variancia total en sus fuentes de variación y grados de libertad se presentan en la Tabla 13.8.

Fuentes de variación	Grados de libertad
Repeticiones (amb)	$a(r-1) = 20$
Genotipos	$g-1 = 19$
Ambientes + Genotipo x Ambiente	$g(a-1) = 180$
Ambiente (lineal)	1
Genotipo x Ambiente (lineal)	$g-1 = 19$
Desvíos de la regresión	$g(a-2) = 160$
Genotipo 1	$a-2 = 8$
Genotipo 2	$a-2 = 8$
...	
Genotipo 20	$a-2 = 8$
Error experimental	$a(g-1)(r-1) = 380$

Tabla 13.8. Esquema de partición del análisis de variancia y de la interacción genotipo x ambiente aplicando al modelo de Eberhart y Russell (1966).

Otra forma de analizar la interacción genotipo x ambiente es descomponer la suma de cuadrados de interacción en tantas sumas de cuadrados como genotipos participan de la evaluación (Wricke y Weber, 1986; Shukla, 1972).

Los análisis multivariados de la interacción genotipo x ambiente consideran a los valores fenotípicos observados en cada ambiente como caracteres diferentes uno de otro. La suma de cuadrados de la interacción se descompone en un número de ejes o componentes principales. Cada eje es ortogonal con los restantes y explica cierto porcentaje de la variancia total de la matriz de g genotipos y a ambientes. Un modelo multivariado de interacción genotipo x ambiente muy empleado es el de efectos principales aditivos y efectos de interacción multiplicativos (modelos AMMI) (Gauch y Zobel, 1996; Crossa *et al.*, 1999; Vargas Hernández y Crossa, 2000).

Los valores observados en la k-ésima repetición del genotipo g-ésimo en el ambiente j-ésimo según el modelo AMMI es (Gauch, 1993):

$$Y_{ijk} = \mu + g_i + a_j + \sum_n \lambda_n Y_{in} \delta_{jn} + R_{ij} + e_{ijk}$$

donde:

μ es la media general,

g_i es el efecto centrado en la media general del genotipo i-ésimo, de modo que $\sum g_i = 0$;

a_j es el efecto centrado en la media general del ambiente j-ésimo, de modo que $\sum a_j = 0$;

λ_n es el autovalor para el n-ésimo eje o componente principal de la interacción;

Y_{in} es el elemento del autovector correspondiente al i-ésimo genotipo en el n-ésimo eje;

δ_{jn} es el elemento del autovector correspondiente al j-ésimo ambiente en el n-ésimo eje;

R_{ij} es el residual del modelo correspondiente al i-ésimo genotipo en interacción con el j-ésimo ambiente; y

e_{ijk} es el error asociado a la r-ésima observación del genotipo i-ésimo (g_i) en el ambiente j-ésimo (a_j).

El modelo AMMI posee dos componentes, uno aditivo correspondiente a la suma ($\mu + g_i + a_j$) y otro multiplicativo correspondiente a la descomposición de la interacción genotipo por ambiente. El número de ejes en que es posible particionar la suma de cuadrados de la interacción es el valor menor de la comparación entre $(g - 1)$ y $(e - 1)$, mientras que el número de grados de libertad correspondiente a cada uno de los ejes es $(g + e - 2n - 1)$, donde n indica el número del eje en cuestión. Consecuentemente los últimos ejes poseen menores números de grados de libertad (Gollob, 1968).

Los primeros ejes del modelo AMMI tienen la propiedad de captar el verdadero patrón de respuesta diferencial de los genotipos expuestos a diferentes ambientes, mientras que los últimos ejes responden mayormente al "ruido" experimental. Por ello y por ser ejes ortogonales (independientes o no correlacionados) entre sí, estos explican porcentajes decrecientes de la variancia fenotípica total. El número de ejes para retener se elige siguiendo un criterio más o menos conservador y parsimonioso. Por lo general se retienen los ejes que acumulan o explican al menos un 70-80 % de la variancia fenotípica total. Frecuentemente alcanza con tener en cuenta los dos primeros ejes y eventualmente puede ser necesario considerar adicionalmente un tercer

eje. El análisis AMMI se representa en un eje de coordenadas, denominado biplot, donde en el eje de las abscisas se representan los puntajes para los genotipos y ambientes correspondientes al primer eje (en realidad los productos $\lambda_1^{0,5} Y_{gi1}$ y $\lambda_1^{0,5} \delta_{aj1}$ para facilitar la interpretación del biplot) y en las ordenadas al promedio fenotípico a través de ambientes. Otro gráfico necesario, cuando se requieren dos ejes para explicar una fracción importante de la variancia de interacción es un biplot donde en el eje de las abscisas se representan los puntajes para los genotipos y ambientes en el primer eje ($\lambda_1^{0,5} Y_{gi1}$ y $\lambda_1^{0,5} \delta_{aj1}$) y en las ordenadas se representan los puntajes para los genotipos y ambientes en el segundo eje ($\lambda_2^{0,5} Y_{gi2}$ y $\lambda_2^{0,5} \delta_{aj2}$). Las interacciones entre genotipos y ambientes que en cada eje posean el mismo signo contribuirán favorablemente al valor fenotípico promedio de ese genotipo en ese ambiente. Los genotipos y ambientes que menor contribución hacen a los efectos de interacción son aquellos que poseen puntajes más próximos a cero para ambos ejes.

Asumiendo que se evaluarán 20 genotipos en diez ambientes y se obtuvieran las gráficas de biplot que se presentan en la Figura 13.2, podemos observar que la descomposición de las medias de genotipos (a través de ambientes) y las de ambientes (a través de genotipos) según el modelo AMMI presentan coordenadas que se distribuyen en los cuatro cuadrantes que quedan delimitados a partir del punto de origen de cada eje. Los ambientes y genotipos que resultan más próximos al valor 0 para ambos ejes, y destacados en color verde son los que menor contribución hacen a la existencia de interacciones genotipo x ambiente. Por el contrario, los genotipos o ambientes más alejados del origen de uno o de ambos ejes son los que mayor contribución hacen a la interacción genotipo por ambiente. La contribución de la interacción genotipo x ambiente al comportamiento promedio de determinado genotipo (i) en determinado ambiente (j) será la sumatoria, través de los ejes que se tengan en cuenta, del producto entre el autovalor y autovector correspondientes:

$$(\lambda_1^{0,5} Y_{gi1} \times \lambda_1^{0,5} \delta_{aj1} + \lambda_2^{0,5} Y_{gi2} \times \lambda_2^{0,5} \delta_{aj2} + \dots + \lambda_n^{0,5} Y_{gin} \times \lambda_n^{0,5} \delta_{ajn})$$

Cuando coinciden los signos de Y_{gin} y de δ_{ajn} la interacción genotipo x ambiente contribuirá positivamente al comportamiento promedio.

La Figura 13.3 corresponde al biplot rendimiento vs. el primer eje AMMI. Claramente, los dos genotipos indicados en verde en la parte superior de los cuadrantes I y IV serían los más promisorios por cuanto reúnen la ventaja de alto rendimiento promedio a través de ambientes y escasa interacción genotipo x ambiente para el eje AMMI que aporta mayor variabilidad a la variancia de interacción. Por el contrario el genotipo identificado en color azul, si bien presentó alto rendimiento promedio y baja interacción genotipo x ambiente explicada por el primer eje, posee elevada interacción a nivel del segundo eje (Figura 13.2), lo cual lo sitúa en un nivel de desventaja respecto de los dos anteriores.

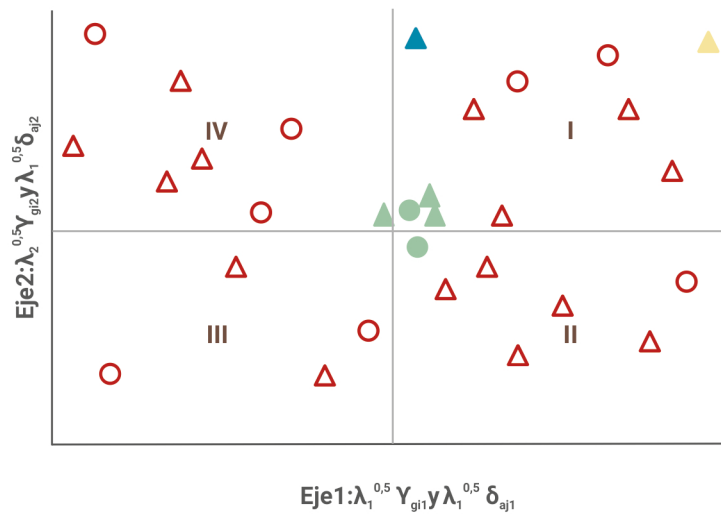


Figura 13.2.

Representación biplot del análisis AMMI de veinte genotipos (triángulos) evaluados en diez ambientes (círculos) correspondiente al primer y segundo eje principal. Los genotipos y ambientes representados en verde son los que menores contribuciones hacen a la variancia de la interacción genotipo x ambiente para ambos ejes. El genotipo indicado en amarillo es el que mayor contribución hace la variancia de interacción genotipo x ambiente para ambos ejes. El genotipo indicado en azul contribuye a la interacción a través del segundo eje.

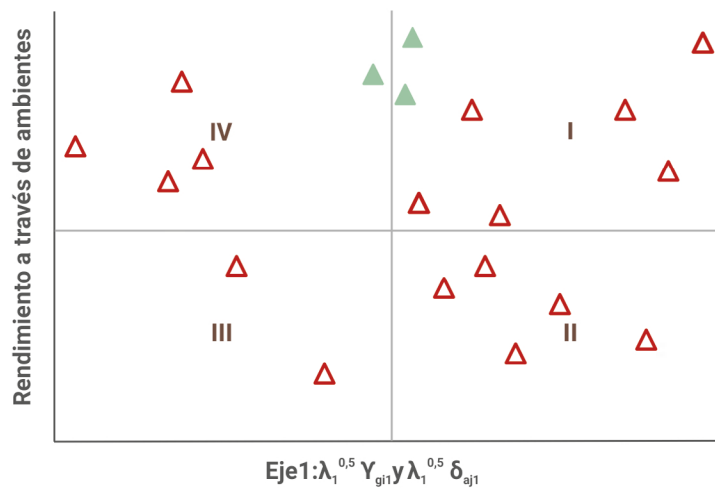


Figura 13.3.

Representación biplot del análisis AMMI de veinte genotipos (triángulos) evaluados en diez ambientes correspondiente al primer eje principal en las abscisas y el rendimiento promedio a través de ambientes en el eje de las ordenadas. Los genotipos y ambientes representados en verde en los cuadrantes I y IV presentaron elevados rendimientos promedio a través de ambientes y escasa contribución a la variancia de interacción.

El autovalor λ de cada eje se relaciona con su contribución a la suma de cuadrados de la interacción genotipo x ambiente IGA, de manera tal que

$$\sum_n \lambda_n^2 = SC (IGA)$$

En el caso de que se emplee un modelo reducido donde se retiene un número menor de ejes (N) respecto del total de ellos aparecerá un residual, de modo que en esos casos

$$\sum_N \lambda_N^2 + \sum_i \sum_j R_{ij}^2 = SC \text{ (IGA)}$$

Los valores del autovector para cada eje de la interacción genotipo x ambiente para los genotipos y para los ambientes están escalados de manera que:

$$\sum_i Y_{gi}^2 = \sum_j \delta_{ej}^2 = 1$$

Otros esquemas de análisis multivariado son los denominados modelos de regresión de sitio (SREG) (Crossa y Cornelius, 2002). En ellos podemos identificar distintas alternativas, dependiendo de las modificaciones que se hagan al modelo básico ya presentado:


$$P_{ijk} = u + a_j + b_{j(k)} + g_i + ga_{ij} + e_{ijk}$$

En la variante GGE (Yan y Hunt, 2002), los términos $g_i + ga_{ij}$ son considerados en forma conjunta y la suma de cuadrados de la interacción se descompone en ejes de modo análogo al modelo AMMI. Permite una representación biplot en un eje de coordenadas y una interpretación geométrica para determinar la cantidad de megambientes y el genotipo ganador en cada uno de estos (Yan y Hunt, 2002).

13.6. Adaptabilidad, estabilidad y confiabilidad genotípicas

La adaptabilidad es la capacidad de un genotipo de lograr buen comportamiento en un rango de ambientes. Para algunos autores, la adaptabilidad puede definirse como la consistencia de buen comportamiento en un rango geográfico y se expresa por el comportamiento promedio. Un genotipo que posee adaptabilidad amplia tendrá un comportamiento previsible frente a la variabilidad ambiental inherente de una región. Un genotipo que posee adaptabilidad específica será aquel que muestra ventajas en su comportamiento frente a otros genotipos en situaciones ambientales más determinadas o menos generales (Lin y Binns, 1994). Como ejemplo de adaptabilidad específica, podemos hablar de genotipos de buen comportamiento en una región donde existe una enfermedad endémica, o variedades de trigo adaptadas a una determinada subregión de cultivo, definida por una serie de factores, como por ejemplo la localización geográfica, fecha de siembra, prácticas de fertilización, etc.

La estabilidad es la capacidad de un genotipo de comportarse consistentemente bien a través de un rango de ambientes. Esa consistencia en el comportamiento tiene que verificarse no solamente a escala temporal, sino también



espacial. A diferencia de la adaptabilidad, que está determinada por la comparación de las medias fenotípicas de los genotipos, la estabilidad que interesa en el fitomejoramiento está asociada a la variabilidad en el comportamiento fenotípico de cada genotipo en comparación con los restantes. Kang (2002) distingue dos criterios para definir estabilidad. La estabilidad estática lo es en sentido biológico, cuando el comportamiento del genotipo es prácticamente constante a través de un rango de calidad ambiental. Los parámetros que se utilizan para estimar la estabilidad estática son independientes del comportamiento de otros genotipos que compartan la evaluación. La estabilidad dinámica, por el contrario, lo es en sentido agronómico y los parámetros que cuantifican la estabilidad de un genotipo o variedad son relativos o comparativos con respecto al grupo de genotipos que comparte la evaluación. Para el estudio de la estabilidad de los genotipos se dispone de métodos tanto paramétricos como no paramétricos (Bujak *et al.*, 2014).

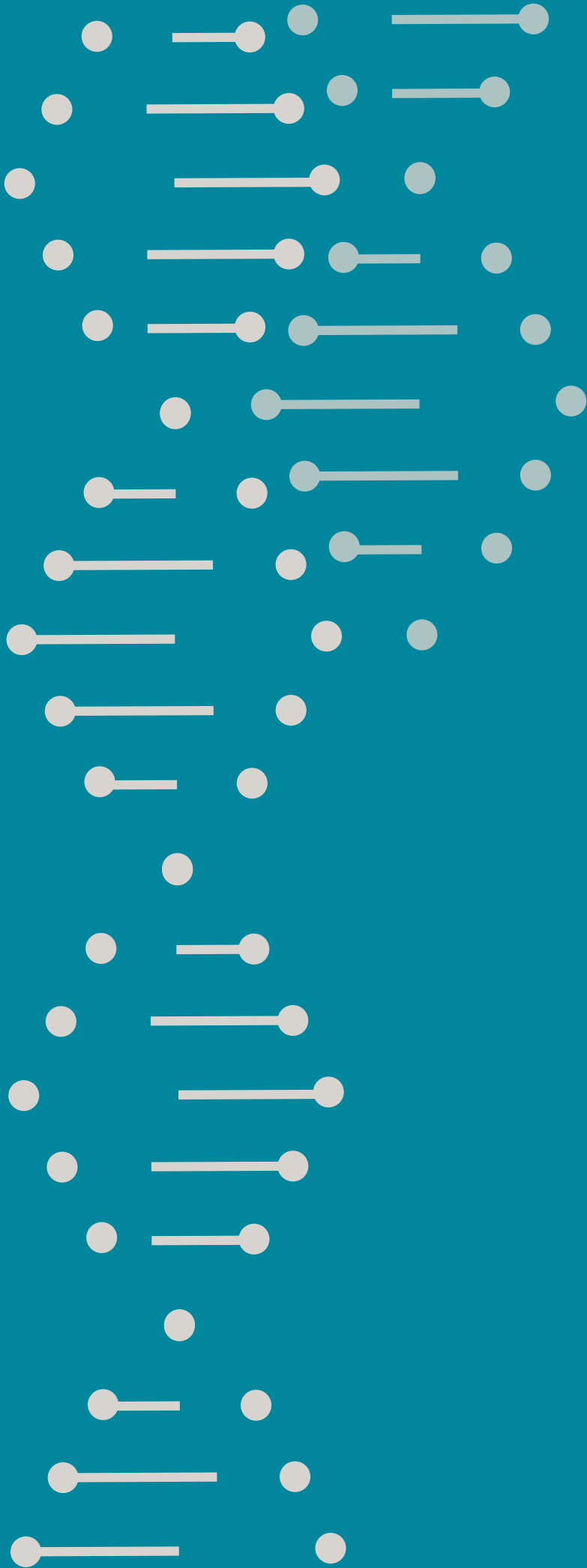
La confiabilidad es la habilidad de un genotipo de combinar buen comportamiento promedio en un determinado rango ambiental y hacerlo de manera estable en el sentido agronómico. Como se ve, la confiabilidad toma en cuenta tanto la media como la variancia de los valores fenotípicos que presenta un genotipo en relación con un grupo que comparte la evaluación.

Los análisis de adaptabilidad, estabilidad y confiabilidad se aplican mayoritariamente al carácter rendimiento de grano, de semilla o de forraje. La selección por productividad (rendimiento) y estabilidad es habitual en los programas de mejoramiento y por ello ambos caracteres tienden a estar genéticamente correlacionados. Sin embargo existen trabajos que comprueban que ambas características se comportan como relativamente independientes en el caso de cruzamientos entre germoplasma no mejorado y germoplasma élite (Eyherabide *et al.*, 2016). Desde el punto de vista de la acción génica involucrada, los efectos no aditivos parecen tener una importante gravitación en la determinación de la estabilidad.

Lunezzo de Oliveira *et al.* (2014) propusieron emplear un índice de selección para seleccionar por adaptabilidad y estabilidad en maíz basado en un índice de distancia genotipo-ideotipo, que emplea como insumo al análisis GGE.

Existe una numerosa oferta de programas o software en continuo crecimiento que permiten la aplicación de diversos modelos de análisis de estabilidad e interacción genotipo x ambiente, entre los que se puede citar a Magari y Kang (1997), Burgueño *et al.* (2001), Hussein *et al.* (2000), etc.

Para una revisión exhaustiva de los métodos de análisis de adaptabilidad, estabilidad y confiabilidad se recomienda la lectura, entre otras, de Shukla (1972), Becker (1981), Becker y León (1988), Lin y Binns (1988), Lin *et al.* (1986), Balzarini (2002) y Kang (2002).





Ploidía y herencia cuantitativa

14.1. Introducción

Muchas especies de interés agronómico tienen la particularidad de poseer un múltiplo mayor que dos del número cromosómico básico de un mismo genoma, o de poseer el doble del número cromosómico básico (x) en más de un genoma. Estas condiciones se denominan genéricamente poliploidía.

Las especies poliploides que poseen múltiplos del número básico cromosómico del mismo genoma se denominan autopoliploides y se clasifican según el número de copias del genoma en triploides ($3x$), autotetraploides ($4x$), autopentaploides ($5x$), etc. Se originan de especies diploides que han multiplicado su número cromosómico básico. Las situaciones de dosis impar del número cromosómico básico se asocian a problemas de este-

ilidad o fertilidad muy reducida. Por el contrario los autotetraploides, por ejemplo, dan gametas viables diploides. Ejemplos de especies cultivadas autoploides o autopoliploides son la papa (*Solanum tuberosum*, $4x=48$), el pasto ovillo (*Dactylis glomerata*, $4x=28$) y la alfalfa (*Medicago sativa*, $4x=42$), (Lacadena, 1970).

Las especies poliploides que poseen múltiplos del número cromosómico básico, pero de más de un genoma se denominan alopoliploides y se clasifican según el número de copias de los genomas que contenga. Los alopoliploides derivan de híbridos entre especies diploides. Los alohexaploides poseen dos juegos cromosómicos básicos de tres genomas diferentes. Los alotetraploides, dos juegos cromosómicos básicos de dos genomas diferentes. Ejemplo de especies de alopoliploides son el trigo pan (*Triticum aestivum*, $6x=42$, tres genomas), el trigo candeal (*Triticum durum*, $4x=28$, dos genomas), la avena (*Avena sativa*, $6x=42$, tres genomas), el algodón (*Gossypium barbadense*, $4x=52$, dos genomas), etc. En algunas de las especies mencionadas pueden encontrarse otros niveles de ploidía. Los alopoliplodes suelen comportarse en su herencia como diploides ya que sus gametas contienen solo un juego cromosómico básico de cada uno de los genomas que lo componen.

Otra clasificación de los organismos poliploides basada en su comportamiento citogenético y sus implicancias en los programas de mejoramiento es en especies poliploides disómicas y polisómicas (Peloquin, 1981; MacKey, 1970). *Triticum aestivum* es una especie poliploide disómica porque los apareamientos solo ocurren entre cromosomas homólogos, sus loci están en estado homocigota, y eventualmente las interacciones entre alelos ubicados en cromosomas homeólogos podrían contribuir a cierto nivel de heterosis (Peloquin, 1981). El apareamiento entre cromosomas ocurre entre aquellos que son homólogos entre sí. Los cromosomas que contienen similar información pero que pertenecen a genomas diferentes se denominan homeólogos. La imposibilidad de formación de multivalentes por apareamiento entre cromosomas homólogos y homeólogos en trigo pan (alohexaploide) se ha demostrado que está bajo control genético (Wricke y Weber, 1986). *Solanum tuberosum*, en cambio, es una especie autopoliploide clasificable como poliploide polisómica, por cuanto el apareamiento e inter cruzamiento cromosómico puede ocurrir entre dos, tres o cuatro cromosomas homólogos. Existen mayores oportunidades de interacciones intralocus o entre diferentes loci. Los organismos polisómicos o autopoliploides poseen en general un alto grado de heterocigosidad, muchas veces enmascarada por genes dominantes.

Entre las angiospermas, el 35 % de las especies es poliploide (Stebbins, 1938), especialmente las especies herbáceas perennes, en menor medida las herbáceas anuales y finalmente las leñosas. En el caso de las gimnospermas, apenas el 4-5 % de ellas es poliploides (Lacadena, 1970). Una excelente revisión de las aplicaciones de la poliploidía en el mejoramiento vegetal puede encontrarse en Sattler *et al.*, 2016.

14.2. Variabilidad genotípica

La herencia de las especies autoploides es más compleja que la de las diploides, por cuanto cada locus de los múltiples cromosomas puede poseer alelos iguales o distintos.

Considerando un locus de un individuo autotetraploide, su constitución puede ser crecientemente compleja, especialmente si existen más de dos alelos por locus. Cuando hay solo dos alelos (A, a), dependiendo del número de alelos de cada clase, un individuo puede ser clasificado como:

- cuadruplexo (aaaa),
- triplexo (Aaaa),
- duplexo (AAaa),
- simplexo (AAAa) y
- nuliplexo (AAAA).

Cuanto mayor sea el número de alelos por locus, mayor será el número de genotipos posibles para cada locus (Tabla 14.1) (Wricke y Weber, 1986). En el caso de que existiera dominancia incompleta, sería advertible un efecto de dosis en función del número de alelos al estado dominante.

Número de genotipos en función de número de alelos				
Tipo	Número de alelos posibles			
	2	3	4	
Monoalélico	A1A1A1A1	A1A1A1A1	A1A1A1A1	A3A3A3A3
(nuliplexos,	k	A2A2A2A2	A2A2A2A2	A4A4A4A4
tetraplexos)		A3A3A3A3		

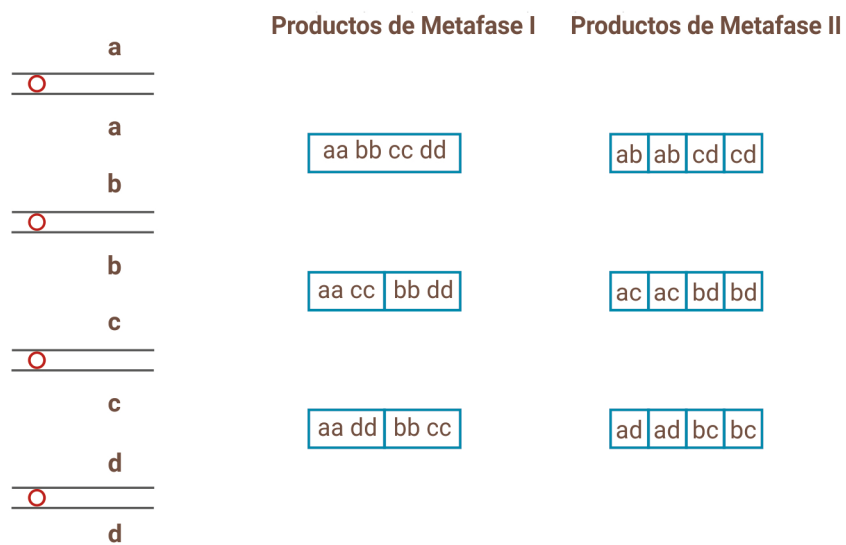
Tabla 14.1. Número de genotipos posibles para un locus de individuos de un organismo autotetraploide (Continúa en la página siguiente)

Tipo	Número de genotipos en función de número de alelos				
	Número de alelos posibles				
	2	3	4		
Bialélico (simplexos y triplexos)	k(k-1)	A1A1A1A2	A1A1A1A2	A1A1A1A2	A2A2A2A1
		A2A2A2A1	A1A1A1A3	A1A1A1A3	A2A2A2A3
		A2A2A2A1	A1A1A1A4	A2A2A2A4	
		A2A2A2A3	A3A3A3A1	A4A4A4A1	
		A3A3A3A1	A3A3A3A2	A4A4A4A2	
		A3A3A3A2	A3A3A3A4	A4A4A4A3	
Trialélico (simplexos y duplexos)	k(k-1)(k-2)/2		A1A2A3A1	A1A1A2A3	A3A3A1A2
			A1A2A3A2	A1A1A2A4	A3A3A1A4
			A1A2A3A3	A1A1A3A4	A3A3A2A4
				A2A2A1A4	A4A4A1A2
				A2A2A3A4	A4A4A2A3
				A2A2A1A3	A4A4A1A3
Cuadrialélico (simplexo)	k(k-1)(k-2)(k-3)/24		A1A2A3A4		

Tabla 14.1.
(Continúa de página anterior)

Podemos clasificar la segregación genética en un autoploiploide en segregación cromosómica y segregación cromatídica (Figura 14.1). En la segregación cromosómica no hay *crossing over*, de modo que en metafase I de la meiosis las dos cromátidas de cada cromosoma se mantienen juntas y en metafase II, cada cromátida del mismo cromosoma termina en gametas diferentes.

[a]



[b]

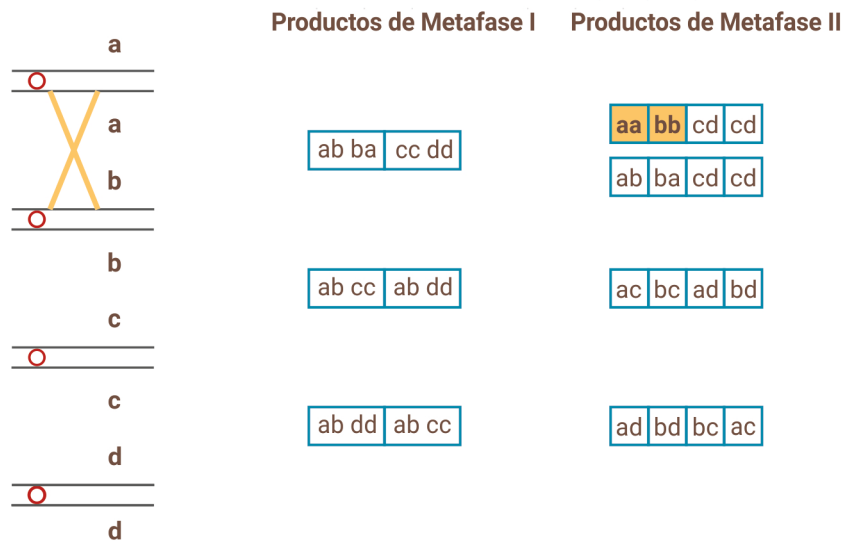


Figura 14.1.

Segregación cromosómica [A] y cromatídica [B] en un individuo autotetraploide. Se considera el caso de un gen con cuatro alelos, identificados como a, b, c y d, todos ellos correspondientes a un locus.

La segregación cromatídica tiene lugar cuando ocurre algún evento de intercambio entre un locus y el centrómero. En tal situación, luego de metafase II las dos cromátidas hermanas de los cromosomas que intervinieron en el entrecruzamiento pueden terminar en las mismas gametas, fenómeno denominado doble reducción. Esas gametas llevan entonces alelos idénticos por descendencia (ver Capítulo 6). La probabilidad de que dos cromátidas hermanas terminen en la misma gameta por doble reducción es α . En un organismo autotetraploide existen 8 cromátidas (2 por cada cromosoma) que se distribuyen al azar y solo 1 de ellas es hermana de cualquiera de las restantes. Por lo tanto $\alpha=1/7$ en el caso de segregación cromatídica completa y $\alpha=0$ cuando hay segregación cromosómica completa, ya que no es posible la doble reducción.

La segregación cromosómica predomina en el caso de genes ubicados más próximos a los centrómeros, ya que su cercanía disminuye las probabilidades de entrecruzamiento entre cromátidas hermanas. La segregación cromatídica, en cambio, es de ocurrencia más probable cuanto más distanciado está el gen del centrómero.

Cuando se analiza la segregación de un carácter de un individuo autoploide, debe considerarse el tipo de segregación y la constitución alélica de los cromosomas homólogos para todos los loci involucrados en el control del carácter.

14.3. Poliploidía y equilibrio poblacional de Hardy-Weinberg

Hemos señalado en capítulos previos que en organismos diploides ($2n=2x$) y bajo apareamiento aleatorio, y en ausencia de migración, mutación y selección, el equilibrio poblacional de Hardy-Weinberg para un locus se alcanza en una generación de apareamiento aleatorio. Ello no es así en el caso de especies autoploides.

En una población de una especie autotetraploide ($2n=4x$), en la cual solo hay dos alelos presentes, A_1 y A_2 , con frecuencias génicas p y q , respectivamente, las combinaciones de gametas diploides darán origen a las cigotas tetraploides conforme se presenta en la Tabla 14.2.



Clases gaméticas			
Clases gaméticas	A1A1	A1A2	A2A2
Clases cigóticas; frecuencias genotípicas			
A1A1	A1A1 A1A1 ; p ⁴	A1A2 A1A1 ; 2p ³ q	A2A2 A1A1 ; p ² q ²
A1A2	A1A1 A1A2; 2p ³ q	A1A2 A1A2 ; 4p ² q ²	A2A2 A1A2; 2pq ³
A2A2	A1A1 A2A2 ; p ² q ²	A1A2 A2A2 ; 2pq ³	A2A2 A2A2 ; q ⁴

Tabla 14.2.

Clases y frecuencias genotípicas luego de una generación de apareamiento aleatorio entre individuos autotetraploides digénicos, con frecuencias génicas p y q, bajo segregación cromosómica.

En la condición de equilibrio, las frecuencias gaméticas serán p² (A₁A₁), 2pq (A₁A₂) y q² (A₂A₂). En cada generación de apareamiento aleatorio, la reducción del desequilibrio es diferente según consideremos segregación cromosómica y segregación cromatídica. En el primer caso la reducción en el desequilibrio de las frecuencias genotípicas respecto de las de Hardy-Weinberg por cada generación de apareamiento aleatorio es de 1/3 y en el segundo caso de 2/7. En el caso general, de que la probabilidad de doble reducción sea mayor que cero y menor de 1/7, la reducción en el desequilibrio será:

$$\Delta = (1 - \alpha)/3$$

Bajo segregación cromosómica, las frecuencias gaméticas y genotípicas en el equilibrio son, respectivamente:

$$(pA_1 + qA_2)^2$$

y

$$(pA_1 + qA_2)^4$$

Tratándose de k alelos, las expresiones pasan a ser, respectivamente:

$$(\sum p_i A_i)^2 \text{ y } (\sum p_i A_i)^4$$

donde i hace referencia al i-ésimo alelo de la serie alélica disponible para el locus A y los exponentes 2 y 4 son el número de genomas en las gametas y el esporofito de un organismo autotetraploide.

La última de esas expresiones puede reescribirse como

$$(p_1A_1 + p_2A_2 + p_3A_3 + \dots p_nA_n)^4$$

donde el número de alelos es n y se trata de un organismo autotetraploide.

Las frecuencias genotípicas en el equilibrio de Hardy-Weinberg para el caso de un locus con segregación cromosómica se presentan en la Tabla 14.3.

Clases genotípicas	Frecuencias genotípicas de equilibrio
$A_1A_1A_1A_1$	p^4
$A_2A_2A_2A_2$	q^4
$A_1A_1A_2A_2$	$6p^2q^2$
$A_1A_1A_1A_2$	$4p^3q$
$A_1A_2A_2A_2$	$4pq^3$

Tabla 14.3.

Clases y frecuencias genotípicas en el equilibrio poblacional luego de varias generaciones de apareamiento aleatorio entre individuos tetraploides digénicos en el caso de segregación cromosómica

Para el caso general, en que la frecuencia de doble reducción está entre 0 y 1/7, las frecuencias gaméticas en el equilibrio poblacional serán (Wricke y Weber, 1986):

$$A_1A_1: p^2 + 3\alpha pq/(2+\alpha)$$

$$A_1A_2: (4-4\alpha)pq/(2+\alpha)$$

$$A_2A_2: q^2 + 3\alpha pq/(2+\alpha)$$

La aproximación a las frecuencias gaméticas y genotípicas de equilibrio demandan, como se indicó más arriba, más de una generación de apareamiento aleatorio.

14.4. Modelo genético

En especies aloploidoides, como por ejemplo el trigo pan, el trigo candeal, el tabaco, que presentan en la meiosis apareamientos cromosómicos bivalentes y herencia disómica, son aplicables los mismos modelos descritos en este libro para los organismos diploides. En otros casos la modelación de su comportamiento hereditario es mucho más complejo y frecuentemente de escaso valor práctico (Kearsey y Pooni, 1996).



La expresión que describe el valor genotípico en una especie autotetraploide, por ejemplo en papa ($2n=4x$), posee más términos causales que en especies diploides:

$$g_{AiAjAkAl} = \alpha_i + \alpha_j + \alpha_k + \alpha_l + \delta_{ij} + \delta_{ik} + \delta_{il} + \delta_{jk} + \delta_{jl} + \delta_{kl} + \gamma_{ijk} + \gamma_{ijl} + \gamma_{ikl} + \gamma_{jkl} + \lambda_{ijkl}$$

donde

$\alpha_i, \alpha_j, \alpha_k, \alpha_l$ son los efectos promedios de los correspondientes alelos del locus A que lleva cada cromosoma,

$\delta_{ij}, \delta_{ik}, \delta_{il}, \delta_{jk}, \delta_{jl}, \delta_{kl}$ son las interacciones bialélicas,

$\gamma_{ijk}, \gamma_{ijl}, \gamma_{ikl}, \gamma_{jkl}$ son las interacciones trialélicas,

λ_{ijkl} es la interacción cuadrialélica.

Los valores genotípicos para individuos que poseen otras estructuras alélicas serán

$$\begin{aligned} g_{AiAiAiAi} &= \alpha_i + \alpha_i + \alpha_i + \alpha_i + \delta_{ii} + \delta_{ii} + \delta_{ii} + \delta_{ii} + \delta_{ii} + \delta_{ii} + \gamma_{iii} + \gamma_{iii} + \gamma_{iii} + \gamma_{iii} + \lambda_{iiii} \\ &= 4\alpha_i + 6\delta_{ii} + 4\gamma_{iii} + \lambda_{iiii} \end{aligned}$$

$$g_{AiAiAiAj} = 3\alpha_i + \alpha_j + 3\delta_{ii} + 3\delta_{ij} + \gamma_{iii} + 3\gamma_{ijj} + \lambda_{iiij}$$

$$g_{AiAiAjAj} = 2\alpha_i + 2\alpha_j + \delta_{ii} + \delta_{jj} + 4\delta_{ij} + 2\gamma_{ijj} + 2\gamma_{ijj} + \lambda_{iiij}$$

$$g_{AiAiAjAk} = 2\alpha_i + \alpha_j + \alpha_k + \delta_{ii} + 2\delta_{ij} + 2\delta_{ik} + \delta_{jk} + \gamma_{ijj} + \gamma_{iik} + 2\gamma_{ijk} + \lambda_{iijk}$$

Estos valores genotípicos pueden expresarse alternativamente como la suma de los valores aditivos y los desvíos de dominancia bialélicos (D), trialélicos (T) y cuadrialélicos (Q):

$$g_{AiAiAiAi} = A$$

$$g_{AiAiAiAj} = A + D_{ij}$$

$$g_{AiAiAjAj} = A + D_{ij}$$

$$g_{AiAiAjAk} = A + D_{ij} + D_{ik} + D_{jk} + T_{ijk}$$

$$g_{AiAjAkAl} = A + D_{ij} + D_{ik} + D_{il} + D_{jk} + D_{jl} + D_{kl} + T_{ijk} + T_{ijl} + T_{jkl} + Q_{ijkl}$$

La media poblacional resultará de la contribución de los individuos homocigotas y de los heterocigotas. Estos últimos, al existir mayor número cromosómico, admiten más clases que en los organismos diploides. Consecuentemente:

$$M = 2(p-q)a + 4p^3qd_3 + 6p^2q^2d_2 + 4pq^3d_1$$

donde la contribución de los individuos homocigotas a la media poblacional es $2(p-q)a$ y la de los individuos heterocigotas d_1 , d_2 y d_3 son los desvíos de dominancia de $A_1A_2A_2A_2$, $A_1A_1A_2A_2$ y $A_1A_1A_1A_2$, respectivamente.

El valor de mejora será la suma de los efectos medios de los alelos que posee el individuo. Si bien al tener gametas diploides las interacciones se transmiten sin cambio a la próxima generación, nada garantiza que puedan transmitirse a las subsiguientes debido a que por efecto de la recombinación genética esas combinaciones podrían perderse. Es una situación análoga a la que puede ocurrir con la epistasis aditiva x aditiva en especies diploides.

Los efectos promedio de A_1 y A_2 serán (Wricke y Weber, 1986):

$$\alpha_{A_1} = (2p-q)a + q^3d_1 + 3pq^2d_2 + 3p^2qd_3$$

$$\alpha_{A_2} = (p-2q)a + p^3d_3 + 3pq^2d_1 + 3p^2qd_2$$

Consecuentemente, el efecto de sustitución será

$$\alpha = \alpha_{A_1} - \alpha_{A_2} = a + p^2(3q-p)d_3 + 3pq(q-p)d_2 + q^2(q-3p)d_1$$

14.5. Variancias y covariancias en especies autoploidoides

La variabilidad de los valores genotípicos de los individuos de la población genera la variancia genotípica, cuya composición causal es la siguiente:

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_T^2 + \sigma_Q^2$$

donde

σ_A^2 es la variancia aditiva,

σ_D^2 es la variancia dialélica debida a la variabilidad entre interacciones dialélicas (δ_{ij}),

σ_T^2 es la variancia trialélica debida a la variabilidad entre interacciones trialélicas (γ_{ijk}),

σ_Q^2 es la variancia cuadrialélica debida a la variabilidad entre interacciones cuadrialélicas (λ_{ijkl}).

Cada una de ellas expresada en función de las frecuencias y efectos génicos son:

$$\sigma_A^2 = 4pq\alpha^2$$

$$\sigma_D^2 = 6p^2q^2\delta^2$$

$$\sigma_T^2 = 4p^3q^3\gamma^2$$

$$\sigma_Q^2 = p^4q^4\lambda^2$$

La Tabla 14.4 presenta los coeficientes que acompañan a cada uno de los componentes de la covariancia entre parientes.

A diferencia de lo que ocurre en especies diploides, (o aloploidos que se comportan meióticamente como tales), las estimaciones que se obtienen de las variancias aditiva y de dominancia a partir de diseños genéticos como los presentados en el capítulo correspondiente a los casos de herencia diploide, estarán sobreestimadas dependiendo de la importancia de las interacciones entre alelos y las epistáticas. Además, en contraposición a las especies diploides, hay interacciones entre alelos del mismo locus que pueden transmitirse a la siguiente generación ya que las gametas contendrán dos juegos cromosómicos básicos, en el caso de los autotetraploides y tres en el caso de los autohexaploides.

Parentesco	σ_A^2	σ_D^2	σ_T^2	σ_Q^2	σ_{AA}^2	σ_{AD}^2	σ_{AT}^2	σ_{AQ}^2	σ_{DD}^2	σ_{DT}^2	σ_{DQ}^2	σ_{TQ}^2
Progenie-Progenitor	1/2	1/6	--	--	1/4	1/12	--	--	1/36	--	--	--
Medios hermanos	1/4	1/36	--	--	1/16	1/144	--	--	1/1296	--	--	--
Hermanos completos	1/2	2/9	1/12	1/36	1/4	1/9	1/24	1/72	4/81	1/54	1/162	1/432

Tabla 14.4. Componentes de la covariancia genotípica entre parientes en una población de una especie autotetraploide. Basado en Wricke y Weber, 1986.

En papa se ha determinado que el rendimiento está asociado a la acción génica sobredominante, la cual surge de la importancia de las interacciones entre alelos. En tal sentido, los loci que poseen constituciones cuadrialélicas son los que permitirían una mayor cantidad de interacciones intralocus (Mendoza y Hayes, 1974).

14.6. Poliploidía y endogamia

Del mismo modo que en las especies diploides, la endogamia disminuye a nivel poblacional la proporción de individuos heterocigotas respecto del total, así como la heterocigosidad a nivel de individuos.

En el caso de autofecundar individuos de un organismo autotetraploide (por ejemplo alfalfa), las familias S_1 tendrán una composición variable en función de la estructura del individuo que se autofecunda (Tabla 14.5).

Progenitor	Composición de la filial S_1				
	AAAA	AAAB	AABB	AABC	ABCD
	(monoalélico)	(bialélico)	(bialélico)	(trialélico)	(tetraalélico)
AAAA	1				
AAAB	1/4	1/2	1/4		
AABB	1/18	4/9	1/2		
AABC	1/36	2/9	1/4	1/2	
ABCD			1/6	2/3	1/6

Tabla 14.5. Composición de la filial S_1 resultante de la autofecundación de individuos tetraploides con diferente estructura alélica.

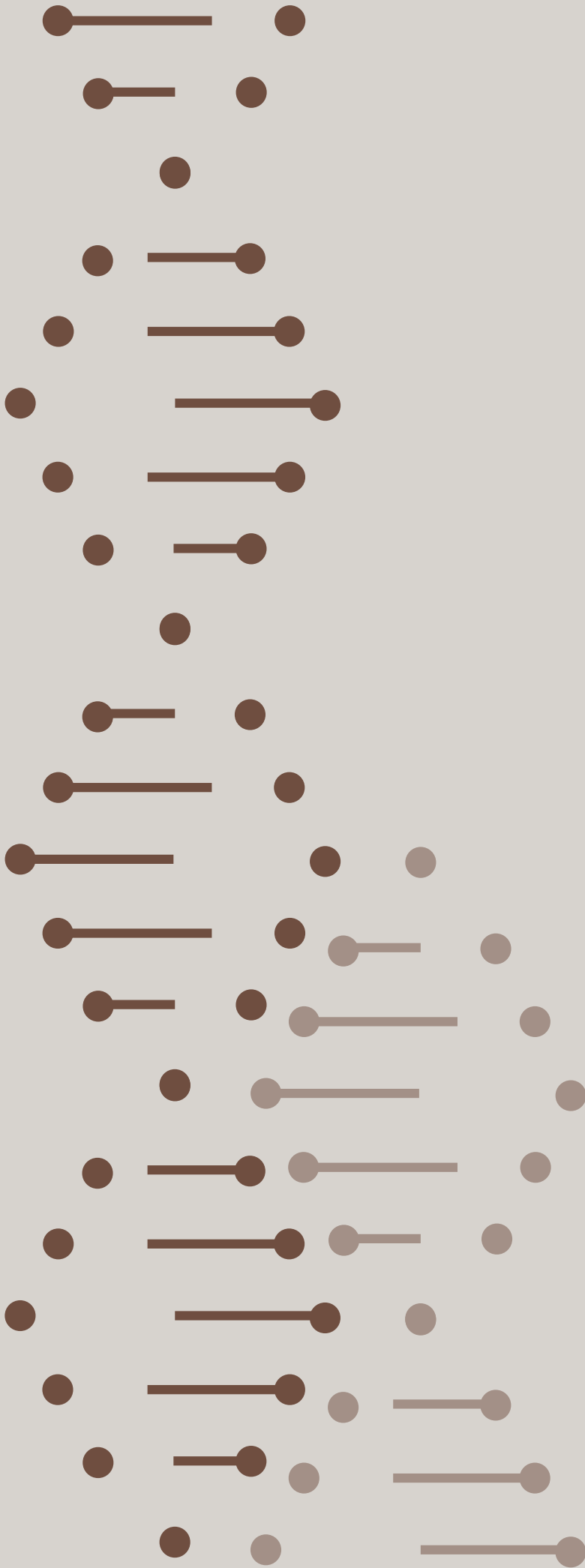
En especies autotetraploides como papa o alfalfa, la descendencia de un individuo puede resultar ser endocriada si sus progenitores estaban relacionados o eran a su vez endocriados. Pero la naturaleza diploide de las gametas hace que, aun en el caso de que los progenitores no estén emparentados, su descendencia tenga un tercio del coeficiente de endogamia promedio de sus padres. Para que la descendencia no posea endocría, debe resultar del cruzamiento entre progenitores no relacionados y sin endocría ($F=0$) (Busbice *et al.*, 1972). Suponiendo la autofecundación de un individuo cuadrialélico $A_i A_j A_k A_l$, este generará tres clases genotípicas con los valores genotípicos, frecuencias y coeficiente de endogamia que se indican en la Tabla 14.6. Nótese que una generación de autofecundación de un individuo poliploide produce un coeficiente de endogamia bastante menor con respecto al caso de la autofecundación de un organismo diploide.

Clases genotípicas	Frecuencia	Valor genotípico	Coefficiente de endogamia (F)
$A_i A_i A_j A_j$	1/6	$2a_i + 2a_j + 4\delta_{ij}$	1/3
$A_i A_i A_j A_k$	2/3	$2a_i + a_j + a_k + 2\delta_{ij} + 2\delta_{ik} + 2\delta_{jk} + \gamma_{ijk}$	1/6
$A_i A_j A_k A_l$	1/6	$a_i + a_j + a_k + a_l + \delta_{ij} + \delta_{ik} + \delta_{il} + \delta_{jk} + \delta_{jl} + \delta_{kl} + \gamma_{ijk} + \gamma_{ijl} + \gamma_{ikl} + \gamma_{jkl} + \lambda_{ijkl}$	0
Promedio de la generación			1/6

Tabla 14.6.

Clases, frecuencias, valores genotípicos y coeficientes de endogamia resultantes de la autofecundación de un individuo autotetraploide cuadrilélico.

Las especies autotetraploides pueden experimentar una importante depresión por endocría, probablemente por la acumulación de genes deletéreos recesivos que se transmiten protegidos en gametas diploides (Busbice *et al.*, 1972).



BIBLIOGRAFÍA

- ACOSTA-PECH, R.; CROSSA, J.; DE LOS CAMPOS, G.; TEYSSÉDRE, S.; CLAUSTRES, B.; PÉREZ-ELIZALDE, S.; PÉREZ-RODRIGUEZ, P. (2017). *Genomic models with genotype x environment interaction for predicting hybrid performance: an application in maize hybrids*. Theor. Appl. Genet., 130, pp. 1431-1440. DOI:10.1007/s00122-017-28989-0
- ALBRECHT, T.; WIMMER, V.; AUINGER, H.J.; ERBE, M.; KNAAK, C.; OUZUNOVA, M.; SIMIANER, H.; SCHÖN, C.C. (2011). "Genome-based prediction of testcross values in maize. Theor. Appl. Genet., 123, pp. 339-350.
- ALLISON, J.C.S.; CURNOW, R.W. (1966). *On the choice of tester parent for the breeding of synthetic varieties of maize (Zea mays L.)*. Crop Sci., 6, pp. 541-544.
- AZZALINI, A.; COX, D.R. (1984). *Two new tests associated with analysis of variance*. J.R. Statist. Soc. B., 46, pp. 335-343.
- BAILEY, T.B. (1977). *Selection limits in self-fertilizing populations following the cross of homozygous lines*. En: Proceedings of the International Conference on Quantitative Genetics. Agosto 16 al 21 de 1976. Editores: Pollak, E.; Kempthorne, O.; Bailey Jr., T.B. Iowa State University Press. Ames, Iowa, EE. UU. p. 872.
- BAILEY, T.B.; COMSTOCK, R.E. (1976). *Linkage and the synthesis of better genotypes in self-fertilizing species*. Crop Sci., 16, pp. 363-370.
- BAKER, R.J. (1978). *Issues in diallel analysis*. Crop Sci., 18, pp. 533-536.
- BAKER, R.J. (1986). *Selection indices in plant breeding*. CRC Press. Boca Raton, Florida, Estados Unidos. p. 218.
- BALZARINI, M. (2002). *Applications of mixed models in plant breeding*. En: Quantitative genetics, genomics and plant breeding. CABI Publishing, Nueva York, EE. UU. p. 400.
- BARTLETT, M.S. (1937). *Properties of sufficiency and statistical tests*. Proc. Roy. Soc. A, 160, pp. 268-282.
- BATESON, W. (1902). *Mendel's principles of heredity*. Cambridge Univ. Press. Reino Unido. p. 212.
- BECKER, H.C. (1981). *Correlations among some statistical measures of phenotypic stability*. Euphytica, 30, pp. 835-840.
- BECKER, H.C.; LEÓN, J. (1988). *Stability analysis in plant breeding*. Plant Breeding, 101, pp. 1-23.
- BERNARDO, R. (1990). *Identifying populations useful for improving parents of a single cross based on net transfer of alleles*. Theor. Appl. Genet., 80, pp. 349-352.
- BERNARDO, R. (1992). *Retention of genetically superior lines during early-generation testcrossing of maize*. Crop Sci., 32, pp. 933-937.
- BERNARDO, R. (2001). *Breeding potential of intra-interheterotic group crosses in maize*. Crop Sci., 48, pp. 68-71.

- BERNARDO, R.** (2010). *Breeding for quantitative traits in plants*. 2.^a edición. Stemma Press. Woodbury, Minnesota, Estados Unidos. p. 390.
- BOCA, R.T.; CANTET, R.J.C.** (2004). *Estimation of additive and dominance variance components in maize using REML and relationship matrices*. *Maydica*, 49, pp. 295-302.
- BUJAK, H.; NOWOSAD, K.; WARZECHA, R.** (2014). *Evaluation of maize hybrids stability using parametric and non-parametric methods*. *Maydica*, 59, pp. 170-175.
- BURGUEÑO, J.; CROSSA, J.; VARGAS, M.** (2001). *SAS program for graphing GE and GGE biplots*. Biometrics and Statistics Unit. CIMMYT, Int. Apdo. Postal 6-641. 06600 México (D.F.), México.
- BUSBICE, T.H.** (1970). *Predicting yield of synthetic varieties*. *Crop Sci.*, 10, pp. 265-269.
- BUSBICE, T.H.; HILL, R.R.; CARNAHAN, H.L.** (1972). *Genetics and breeding procedures*. En: *Alfalfa science and technology*. Editor: Hanson, C.H. American Society of Agronomy, Madison, Wisconsin, EE. UU. p. 812.
- CABALLERO, A.** (1994). *Developments in the prediction of effective population size*. *Heredity*, 73, pp. 657-679.
- CHAKRABORTY, R.; MOREAU, L.; DEKKERS, J.C.M.** (2002). *A general method to optimize selection on multiple identified trait loci*. *Genetics, Selection, Evolution*, 34, pp. 145-170.
- CHEVERAUD, J.M.; ROUTMAN, E.J.** (1995). *Epistasis and its contribution to genetic variance components*. *Genetics*, 139, pp. 1455-1461.
- COCHRAN, W.G.** (1951). *Improvement by means of selection*. Proceedings of the second Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley, California, EE. UU. pp 449-470.
- COCKERHAM, C.C.** (1963). *Estimation of genetic variances*. En: *Statistical Genetics and Plant Breeding*. Editor: Hanson, W.D.; Robinson, H.F. North Carolina State College. National Academy of Sciences-National Research Council, Washington, D.C., Estados Unidos. p. 623.
- COCKERHAM, C.C.** (1967). *Prediction of double crosses from single crosses*. *Züchter*, 37, pp. 160-169.
- COMBS, E; BERNARDO, R.** (2013). *Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers*. *The Plant Genome*, 6 (1), pp. 1-7.
- COMPTON, W.A.; COMSTOCK, R.E.** (1976). *More on modified-ear-to-row selection in corn*. *Crop Sci.*, 16, p. 122.
- COMSTOCK, R.E.** (1978). *Quantitative genetics in maize breeding*. En: *Maize Breeding and Genetics*, Ed: Walden, DB. Wiley, J. and Sons, Nueva York, Estados Unidos. p. 794.
- COMSTOCK, R.E.** (1996). *Quantitative genetics with special reference to plant and animal breeding*. Iowa State University Press, Ames, Iowa, Estados Unidos. p. 421.
- COMSTOCK, R.E.; ROBINSON, H.F.** (1948). *The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance*. *Biometrics*, 4, pp. 254-266.
- COMSTOCK, R.E.; ROBINSON, H.F.; HARVEY, P.H.** (1949). *A breeding procedure designed to make maximum use of both general and specific combining ability*. *Agron. J.*, 42, pp. 360-367.
- COX, D.F.** (1983). *Planning, executing and interpreting experiments*. Iowa State University, Ames, Iowa, Estados Unidos.
- CROSBIE, T.M.; MOCK, J.J.; SMITH, O.S.** (1980). *Comparison of gains predicted by several selection methods for cold tolerance traits of two maize populations*. *Crop Sci.*, 20, pp. 649-655.
- CROSSA, J.** (1989). *Methodologies for estimating the sample size required for genetic conservation of outbreeding crops*. *Theor. Appl. Genet.*, 77, pp. 153-161.
- CROSSA, J.; CORNELIUS, P.L.** (2002). *Linear-bilinear models for the analysis of genotype-environment interaction*. En: *Quantitative genetics, genomics, and plant breeding*. Editor: Kang, M.S. CABI Publishing, Nueva York, Estados Unidos. p. 400.
- CROSSA, J.; GAUCH, H.G.; ZOBEL, R.W.** (1990). *Additive main effects and multiplicative interaction analysis of two international maize cultivar trials*. *Crop Sci.*, 30, pp. 493-500.
- CROW, J.F.; KIMURA, M.** (1970). *An introduction to population genetics theory*. Harper y Row, Nueva York, Estados Unidos. p. 656.
- CRUZ, C.D.; REGAZZI, A.J.** (1994). *Modelos biométricos aplicados ao melhoramento genético*. Universidades Federal de Vicosa, Minas Gerais, Brasil. p. 390.

- CUNNINGHAM, E.P. (1969). *The relative efficiencies of selection indexes*. Acta Agricultura Scandinavica, 19, pp. 45-48.
- CUNNINGHAM, E.P.; MOAN, R.A.; GJEDREM, T. (1970). *Restriction of selection indexes*. Biometrics, 26, pp. 67-74.
- CUNNINGHAM, E.P. (1975). *Multi-stage index selection*. Theoretical and Applied Genetics, 46, pp. 55-61.
- DAETWYLER H.D.; VILLANUEVA, B.; WOOLLIAMS, J.A. (2008). *Accuracy of predicting the genetic risk of disease using a genome-wide approach*. PLoS ONE, 3, e3395.
- DARWIN, C. (1889). *The effects of cross and self fertilization in the vegetable kingdom*. Reimpreso por BiblioBazaar, (2015). p. 498.
- DEKKERS, J.C.M. (2007). *Prediction of response to marker –assisted and genomic selection using index theory*. J. Anim. Breed. Genet., 124, pp. 331-341.
- DEKKERS, J.C.M.; VAN ARENDONK, J.A.M. (1998). *Optimizing selection for quantitative traits with information on an identified locus in outbred populations*. Genetical Research, 71, pp. 257-275.
- DE LOS CAMPOS, G.; HICKEY, J.M.; PONG-WONG, R.; DAETWYLER, H.D.; CALUS, M.P.L. (2013). *Whole-genome regression and prediction methods applied to plant and animal breeding*. Genetics, 193, pp. 327-345.
- DE VILMORIN, L. (1850). *Note sur un projet d' experience avant pou bt d'augmenter la richesse saccharine de la betterave*. Bulletin des Séances de la Société Impériale d' Agriculture, 2e série 6, p. 169.
- DICKERSON, G.E. (1963). *Biological interpretation of the genetic parameters of populations*. En: *Statistical Genetics and Plant Breeding*. Editores: Hanson, W.D.; Robinson, H.F. North Carolina State College. National Academy of Sciences-National Research Council, Washington, D. C., EE. UU. p. 643.
- DUDLEY, J.W. (1982). *Theory for transfer of alleles*. Crop Sci. 22, pp. 631-637.
- DUDLEY, J.W. (1984). *Identifying parents for use in a pedigree breeding program*. 39th Annual Corn & Sorghum Research Conference. Am. Seed Trade Assoc. Chicago, Illinois, Estados Unidos. pp. 176-187.
- DUDLEY, J.W. (1984a). *A method for identifying lines for use in improving parents of a single cross*. Crop Sci., 24, pp. 355-357.
- DUDLEY, J.W. (1984b). *A method for identifying populations containing favorable alleles not present in elite germplasm*. Crop Sci., 24, pp. 1053-1054.
- DUDLEY, J.W. (1987). *Modifications of methods for identifying populations to be used for improving parents of elite single crosses*. Crop Sci., 27, pp. 940-943.
- DUDLEY, J.W. (2007). *From means to QTL: The Illinois long term selection experiment as a case study in Quantitative Genetics*. Crop Sci., 47 (S3), pp. S20-S31.
- DUDLEY, J.W.; MOLL, R.H. (1969). *Interpretation and use of estimates of heritability and genetic variances in plant breeding*. Crop Sci., 9, pp. 257-262.
- EAST, E.M. (1910). *A mendelian interpretation of variation that is apparently continuous*. Am. Nat., 44, pp. 65-82.
- EBERHART, S.A. (1964). *Least squares method for comparing progress among recurrent selection methods*. Crop Sci., 4, pp. 230-231.
- EBERHART, S.A. (1964). *Theoretical relations among single, three way, and double cross hybrids*. Biometrics, 20, pp. 522-539.
- EBERHART, S.A.; HARRISON, M.N.; OGADA, F. (1967). *A comprehensive breeding system*. Züchter, 37, pp. 169-174.
- EBERHART, S.A. (1970). *Factors affecting efficiencies of breeding methods*. African Soils, 15, pp. 669-680.
- EBERHART, S.A.; RUSSELL, W.A. (1966). *Stability parameters for comparing varieties*. Crop Sci., 6, pp. 36-40.
- EHDAIE, B.; CRESS, C.E. (1973). *Simulation of cyclic single cross selection*. Theoretical an Applied Genetics, 43, pp. 374-380.
- ELSTON, R.C. (1963). *A weight-free index for the purpose of ranking and selection with respect to several traits at a time*. Biometrics, 19, pp. 85-97.
- EMERSON, R.A. (1910). *The inheritance of sizes and shapes in plants*. A preliminary note. Am. Nat., 44, pp. 739-746.
- EMPIG, L.T.; GARDNER, C.O.; COMPTON, W.A. (1972). *Theoretical gains for different population improvement procedures*. University of Nebraska. College of Agriculture. MP26.

- EWENS, W.J. (1963). *Numerical results and diffusion approximations in a genetic process*. Biometrika, 50, pp. 241-249.
- EYHERABIDE, G.H.; HALLAUER, A.R. (1991a). *Reciprocal full-sib recurrent selection in maize: I: Direct and indirect responses*. Crop Sci., 31, pp. 952-959.
- EYHERABIDE, G.H.; HALLAUER, A.R. (1991b). *Reciprocal full-sib recurrent selection in maize: II: Contributions of additive, dominance and genetic drift effects*. Crop Sci., 31, pp. 1442-1448.
- EYHERABIDE, G.H.; LÓPEZ, C.G.; BOCA, R.T. (2016). *Gene action controlling stability and adaptability in maize*. Maydica, 61, p. 9. (Disponible: <https://journals/crea.4science.it/index.php/maydica/article/view/1547> verificado: agosto de 2020).
- FALCONER, D.S. (1981). *Introduction to Quantitative Genetics*. Longman, Londres, Reino Unido. p. 340.
- FALCONER, D.; MACKAY, T.F. (2001). *Introducción a la genética cuantitativa*. Editorial Acribia S.A., Zaragoza, España. p. 500.
- FEHR, W. (1987). *Principles of cultivar development. Theory and technique*. Volumen 1. Macmillan Publishing Company, Nueva York, EE. UU. p. 536.
- FIGUEIREDO SANTOS, M.; MARINHO CÁMARA, T.M.; VITTI MORO, G.; NASSAU COSTA, E.F.; SOUZA JR., C.L. (2007). *Responses to selection and changes in combining ability after three cycles of a modified reciprocal recurrent selection in maize*. Euphytica, 157, pp. 185-194.
- FINNEY, D.J. (1958). *Plant selection for yield improvement*. Euphytica, 7, pp. 83-106.
- FINNEY, D.J. (1962). *Genetic gains under three methods of selection*. Genet. Res. Camb., 3, pp. 417-423.
- FISHER, R.A. (1918). *The correlation between relatives on the supposition of Mendelian inheritance*. Trans. R. Soc. Edinburgh, 52, pp. 399-433.
- FISHER, R.A. (1922). *On the mathematical foundation of theoretical statistics*. Philos. Trans. Roy. Soc. London Ser. A., 222, pp. 309-368.
- FISHER, R.A. (1936). *The use of multiple measurements in taxonomic problems*. Ann. Eugenics, 7, pp. 179-188.
- FOX, P.N.; ROSIELLE, A.A. (1982). *Reducing the influence of environmental main effects on pattern analysis of plant breeding environments*. Euphytica, 31, pp. 645-656.
- FRANCIS, D.M.; MERK, H.L.; NAMUTH-COVERT, D. (2012). *Gene pyramiding using molecular markers*. Extension. (Disponible: articles.extension.org/pages/32465/gene-pyramiding-using-molecular-markers verificado: agosto de 2020).
- FREEMAN, A. (1988). *Notas de curso AnSci. 651, Department of Animal Science*. Iowa State University.
- GALE, J.S.; LAWRENCE, M.J. (1984). *The decay of variability*. En: *Crop Genetic Resources: Conservation & Evaluation*. Editores: Holden, J.H.W.; Williams, J.T. George Allen & Unwin. Londres.
- GALTON, F.R.S. (1889). *Natural inheritance*. Macmillan and Co. Londres. p. 266.
- GARDNER, C.O. (1961). *An evaluation of the effects of mass selection and seed irradiation with thermal neutrons on yield of corn*. Crop Sci., 1, pp. 241-245.
- GARDNER, C.O. (1967). *Simplified methods for estimating constant and computing sum of squares for a diallel cross analysis*. Fitotecnia Latinoamericana, 4, pp. 1-12.
- GARDNER, C.O.; EBERHART, S.A. (1966). *Statistical genetic theory and procedures useful in studying varieties and intervarietal crosses in maize*. Heterosis in intervarietal crosses of maize. CIMMYT. Research Bulletin N.º 2.
- GARDNER, C.O.; LONNQUIST, J.H. (1966). *Analysis and interpretation of the variety cross diallel and related populations*. Biometrics, 22, pp. 439-452.
- GAUCH, H.G. (1993). *Matmodel Version 2.0: AMMI and related analyses for two-way data matrices*. Microcomputer Power, Ithaca, Nueva York.
- GAUCH, H.G.; ZOBEL, R.W. (1996). *AMMI analysis of yield trials*. En: *Genotype-by-environment interaction*. Editores: Kang, M.S.; Gauch, H.G. CRC Press, Boca Ratón, Florida, EE. UU. pp. 85-122.
- GAUT, B.S.; LONG, A.D. (2003). *The lowdown on linkage disequilibrium*. The Plant Cell, 15, pp. 1502-1506.
- GERLOFF, J.E.; SMITH, O.S. (1988). *Choice of method for identifying germplasm with superior alleles. 1: Theoretical results*. Theor. Appl. Genet., 76, pp. 209-216.
- GOLLOB, H.F. (1968). *A statistical model which combines features of factor analytic and analysis of variance techniques*. Psychometrika, 33, pp. 73-115.

- GOODNIGHT, C.J.** (1999). *Epistasis and heterosis*. En: Genetics and Exploitation of heterosis in crops. Editores: Coors, J.G.; Pandey, S. ASA-CSSA, Madison, Wisconsin, EE. UU. p. 524.
- GRAFIUS, J.E.** (1965). *A geometry of plant breeding*. Michigan State University Agricultural Experiment Station Res. Bull. N.º 7, East Lansing. p. 59.
- GRIFFING, B.** (1956). *Concept of general and specific combining ability in relation to diallel crossing systems*. Australian J. Biol. Sci., 9, pp. 465-493.
- GUPTA, P.K.; RUSTGI, S.; KULWAL, P.L.** (2005). *Linkage disequilibrium and association studies in higher plants: Present status and future prospects*. Plant Molecular Biology, 57, pp. 461-485.
- HALLAUER, A.R.** (1967). *Development of single-cross hybrids from two-eared maize populations*. Crop Sci., 7, pp. 192-195.
- HALLAUER, A.R.** (1973). *Hybrid development and population improvement in maize by reciprocal full-sib selection*. Egypt. J. Genet. Cytol., 2, pp. 84-101.
- HALLAUER, A.R.** (1975). *Relation of gene action and type of tester in maize breeding procedures*. 30th Annual Corn & Sorghum Research Conference. American Seed Trade Association. ASTA. pp. 150-164.
- HALLAUER, A.R.** (1984). *Reciprocal full-sib selection in maize*. Crop Sci., 24, pp. 755-759.
- HALLAUER, A.R.** (1987). *Compendium of recurrent selection methods and their application*. CRC Critical Reviews in Plant Sciences, 3, pp. 1-34.
- HALLAUER, A.R.** (1992). *Breeding systems. CRC Handbook of Plant Science in Agriculture*, Volumen 1, pp. 61-87.
- HALLAUER, A.R.** (1999). *Temperate maize and heterosis*. En: Genetics and Exploitation of heterosis in crops. Editores: Coors, J.C.; Pandey, S. ASA-CSSA, Madison, Wisconsin, EE. UU. p. 524.
- HALLAUER, A.R.; EBERHART, S.A.** (1970). *Reciprocal full-sib selection*. Crop Sci., 10, pp. 315-316.
- HALLAUER, A.R.; MIRANDA FILHO, J.B.** (1988). *Quantitative genetics in maize breeding*. Iowa State University Press, 2.a edición. Ames, Iowa, EE. UU. p. 468.
- HALLAUER, A.R.; CARENA, M.J.; MIRANDA FILHO, J.B.** (2010). *Handbook of plant breeding*. Vol. 6. Quantitative genetics in maize breeding. Springer Science + Business Media, LLC. Nueva York, EE. UU. p. 663.
- HAMMOND, J.J.; GARDNER, C.O.** (1974). *Modification of the variety cross diallel model for evaluating cycles of selection*. Crop Sci., 14, pp. 6-8.
- HANSON, W.D.** (1963). *Heritability*. En: Statistical Genetics and Plant Breeding. Editores: Hanson, W.D.; Robinson, H.F. North Carolina State College. National Academy of Sciences-National Research Council, Washington, D.C.; EE. UU. p. 623.
- HANSON, W.D.** (1983). *Quantification of specific gene interaction effects between populations utilizing diallel information*. Crop Sci., 23, pp. 769-774.
- HANSON, W.D.** (1987). *Evaluating genetic changes associated with selection utilizing information from diallel mating designs*. Crop Sci., 27, pp. 919-923.
- HANSON, R.W.** (1989). *Standard errors for heritability and expected selection response*. Crop Sci., 29, pp. 1561-1562.
- HANSON, W.D.; MOLL, R.H.** (1986). *An analysis of changes in dominance-associated gene effects under intrapopulation and interpopulation selection in maize*. Crop Sci., 26, pp. 268-273.
- HARDY, G.H.** (1908). *Mendelian proportions in a mixed population*. Science, 28, pp. 49-50.
- HARTLEY, H.O.; RAO, J.N.** (1967). *Maximum likelihood estimation for the mixed analysis of variance model*. Biometrika, 54, pp. 93-108.
- HARVILLE, D.A.** (1975). *Index selection with proportionality constraints*. Biometrics, 31, pp. 223-225.
- HAYES, B.J.; DAETWYLER, H.D.; BOWMAN, P.; MOSER, G.; TIER, B.; CRUMP, R.; KHATKAR, M.; RAADSMA, H.W.; GODDARD, M.E.** (2013). *Accuracy of genomic selection: comparing theory and results*. Proc. Assoc. Advmt. Anim. Breed. Genet., 18, pp.34-37.
- HAZEL, L.N.** (1943). *The genetic basis for construction selection indexes*. Genetics, 28, pp. 476-490.
- HAZEL, L.N.; LUSH, J.L.** (1942). *The efficiency of three methods of selection*. J. Hered., 33, pp. 393-399.

- HEFFNER, E.L.; SORRELLS, M.E.; JANNINK, J.L. (2009). *Genomic selection for crop improvement*. Crop Sci., 49, pp. 1-12.
- HELMS, T.C; HAMMOND, J.J. (2006). *Genetic gain equation with correlated genotype x environment effects*. Crop Sci., 46, pp. 1137-1142.
- HENDERSON, C.R. (1963). *Selection index and expected genetic advance*. En: Statistical Genetics and Plant Breeding. Editores: Hanson, W.D.; Robinson, H.F. NAS-NCR (982), EE. UU. p. 623.
- HESLOT, N.; JANNINK, J.L.; SORRELLS, M.E. (2013). *Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data*. Crop Sci., 53, pp. 921-933.
- HO, Y.T.; COMSTOCK, R.E. (1980). *Combining superior alleles from two homozygous populations in a cross-fertilizing species*. Genet. Res. Camb. pp. 277-287.
- HOLLAND, J.B. (2007). *Genetic architecture of complex traits in plants*. Current Opinion in Plant Biology, 10, pp. 156-161.
- HOPKINS, C.G. (1899). *Improvement in the chemical composition of the corn kernel*. Illinois Agric. Exp. Stn. Bull., 55, pp. 205-240.
- HULL, F.H. (1945). *Recurrent selection and specific combining ability in corn*. J. Am. Soc. Agron., 37, pp. 134-145.
- HUSSEIN; M.A.; BJORNSTAD, A.; AASTVEIT, A.H. (2000). *SASG x ESTAB: A SAS program for computing genotype x environment stability statistics*. Agron. J., 92, pp. 454-459.
- IWATA, H.; JANNINK, J.L. (2011). *Accuracy of genomic selection prediction in barley breeding programs: a simulation study based on the real single nucleotide polymorphism data of barley breeding lines*. Crop Sci., 51, pp. 1915-1927.
- JENKINS, M.T. (1934). *Methods of estimating the performance of double crosses in corn*. J. Am. Soc. Agron., 26, pp. 199-204.
- JENKINS, M.T. (1940). *The segregation of genes affecting yield of grain in maize*. J. Am. Soc. Agron., 32, pp. 55-63.
- JOHANSENN, W. (1903). *On heredity in society and in pure lines*. Erbllichkeit in Populationen und in reinen Linien, 3, pp. 247-270.
- JONES, L.P.; COMPTON, W.A.; GARDNER, C.O. (1971). *Comparison of full and half-sib reciprocal recurrent selection*. Theoretical and Applied Genetics, 41, pp. 36-39.
- KANG, M.S. (2002). *Genotype-environment interaction: progress and prospects*. Quantitative genetics, genomics and plant breeding. CABI Publishing, EE. UU. p. 400.
- KEARSEY, M.J.; POONI, H.S. (1996). *The genetical analysis of quantitative traits*. Chapman and Hall. Londres, Reino Unido. pp. 381.
- KELKER, D.; KELKER, H. (1986). *The effect of skewness on selection in a plant breeding program*. Euphytica, 35, pp. 303-309.
- KEMPTHORNE, O. (1954). *The correlation between relatives on the supposition of Mendelian inheritance*. Proc. R. Soc. Londres, B143, pp. 103-113.
- KEMPTHORNE, O. (1957). *An introduction to Genetic Statistics*. Wiley, Nueva York, EE. UU. p. 545.
- KEMPTHORNE, O; NORDSKOG, A.W. (1959). *Restricted selection indices*. Biometrics, 15, pp. 10-19.
- KEERATINIJAKAL, V.; LAMKEY, K.R. (1993). *Responses to reciprocal recurrent selection in BSSS and BSCB1 maize populations*. Crop Sci., 33, pp. 73-77.
- KINMAN; M.L.; SPRAGUE, G.F. (1945). *Relation between the number of parental lines and theoretical performance of synthetic varieties of corn*. J. A. Soc. Agron., 37, pp. 341-351.
- KLEI, B.; TSURUTA, S. (2008). *Approximate variance for heritability estimates*. (Disponible: http://nce.ads.uga.edu/html/projects/AL_SE_revised.pdf verificado: 13 de marzo de 2018).
- KNIGHT, T.A. (1799). *An account of some experiments on the fecundation of vegetables*. Philos. Trans. R. Soc. Londres, 89, pp.195-204.
- KNIGHT, R.L. (1948). *Dictionary of Genetics, including terms used in cytology, animal breeding and evolution*. Waltham Mass. Chronica Botannica Co. EE. UU. p. 183.
- LACADENA, J.R. (1970). *Genética vegetal. Fundamentos de su aplicación*. 2.a edición. Madrid, España. p. 429.

- LAMKEY, K.R.; EDWARDS, J.W. (1999). *Quantitative genetics of heterosis*. En: Genetics and Exploitation of heterosis in crops. Editores: Coors, J.C.; Pandey, S. ASA-CSSA, Madison, Wisconsin, EE. UU. p. 524.
- LAMKEY, K.R.; SCHNICKER, B.S.; MELCHINGER, A.E. (1995). *Epistasis in an elite maize hybrid and choice of generation for inbred line development*. *Crop Sci.*, 35, pp.1272-1281.
- LANDE, R.; THOMPSON, R. (1990). *Efficiency of marker-assisted selection in the improvement of quantitative traits*. *Genetics*, 124, pp. 743-756.
- LERNER, I.M. (1958). *The genetic basis of selection*. Wiley, Nueva York, EE. UU. p. 298.
- LI, C.C. (1975). *Path analysis-a primer*. The Boxwood Press. Pacific Grove, California, EEUU. p. 346.
- LIN, C.S.; BINNS, M.R. (1988). *A superiority measure of cultivar performance for cultivar x location data*. *Can. J. Plant Sci.*, 68, pp.193-198.
- LIN, C.S.; BINNS, M.R. (1994). *Concepts and methods for analyzing regional trial data for cultivar and location selection*. Editor: Janick, J. *Plant Breeding Reviews*, 76, John Wiley & Sons Inc. Nueva York. pp. 271-297.
- LIN, C.S.; BINNS, M.R.; LEFKOVITCH, L.P. (1986). *Stability analysis: where do we stand?* *Crop Sci.*, 26, pp. 894-900.
- LIN, Z.; HAYES, B.J.; DAETWYLER, H.D. (2014). *Genomic selection in crops, trees and forages: a review*. *Crop and Pasture Science*. <https://www.publish.csiro.au/cp/CP13363>
- LONNQUIST, J.H. (1964). *Modification of the ear-to-row procedure for the improvement of maize populations*. *Crop Sci.*, 4, pp. 227-228.
- LONNQUIST, J.H.; WILLIAMS, N.E. (1967). *Development of maize hybrids through selection among full sib families*. *Crop Sci.*, 7, pp. 369-370.
- LUNEZZO DE OLIVEIRA, R.; GARCÍA VON PINHO, R.; FURTADO FERREIRA, D.; MIRANDA PIRES, L.P.; COSTA MELO, W.M. (2014). *Selection index in the study of adaptability and stability in maize*. ID 360570, p. 6. <http://dx.doi.org/10.1155/2014/360570>
- MACKAY, J. (1970). *Significance of mating systems for chromosomes and gametes in polyploids*. *Hereditas*, 66, pp. 165-176.
- MAGARI, R.; KANG, M.S. (1997). *SAS-STABLE: stability analyses of balanced and unbalanced data*. *Agron. J.*, 89, pp. 929-932.
- MAINLAND, G.B. (1923). *Muller's method of calculating population sizes*. *J. Heredity*, 42, pp. 237-240.
- MALECOT, G. (1948). *Les mathématiques de l'hérédité*. Masson et Cie, Paris, Francia. p. 66.
- MARIOTTI, J.A. (1986). *Fundamentos de Genética Biométrica. Aplicaciones al mejoramiento genético vegetal*. Secretaría General de la Organización de Estados Americanos. Monografía N.º 32. p. 152.
- MÁRQUEZ-SÁNCHEZ, F. (1982). *Modifications to cyclic hybridization in maize with single eared plants*. *Crop Sci.*, 22, pp. 314-319.
- MÁRQUEZ-SÁNCHEZ, F. (1987). *A proposed modification to the Paterniani-Vencovsky method of reciprocal recurrent selection*. *Maydica*, 32, pp. 341-343.
- MÁRQUEZ-SÁNCHEZ, F. (2009). *Expected inbreeding in recurrent selection in maize.III: selection in S1 lines and full-sib and half-sib families*. *Maydica*, 54, pp. 109-111.
- MÁRQUEZ-SÁNCHEZ, F. (2013). *Cálculo de la endogamia en la selección recurrente de familias de maíz*. *Revista Mexicana de Ciencias Agrícolas*, 4, pp. 153-158.
- MATHER, K. (1949). *Biometrical genetics. The study of continuous variation*. Methuen, Londres. Gran Bretaña. p. 167.
- MATHER, K.; JINKS, L.J. (1971). *Biometrical genetics*. Cornell Univ. Press. Ithaca, Nueva York, EE. UU.
- MELCHINGER, A.E.; SCHMIDT, W.; GEIGER, H.H. (1988). *Comparison of Testcrosses Produced from F2 and First Backcross Populations in Maize*. *Crop Sci.*, 28, pp. 743-749.
- MENDEL, G.J. (1865). *Experiments on plant hybridization*. Proceedings of the Brünn National Science Society, pp. 3-47.
- MENDOZA, H.A.; HAYNES, F.L. (1973). *Some aspects of breeding and inbreeding in potatoes*. *American Potato Journal*, 50, pp. 216-222.
- MENDOZA, H.A.; HAYNES, F.L. (1974). *Genetic relationship among potato cultivars grown in the United States*. *Hort Science*, 9, pp. 328-330.

- MEUWISSEN, T.H.E. (2009). Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetics, Selection, Evolution*, 41, pp. 35-44. Doi:10.1186/1297-9686-41-35
- MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, pp. 1819-1829.
- MIRANDA-FILHO, J.B. (1982). Mating schemes to predict gain on heterosis in reciprocal recurrent selection. *Rev. Brasil. Genet.*, 4, pp. 745-760.
- MIRANDA-FILHO, J.B.; GERALDI, I.O. (1984). An adapted model for the analysis of partial diallel crosses. *Rev. Brasil. Genet.*, 7, pp. 677-688.
- MIRANDA-FILHO, J.B.; PATERNIANI, E. (1983). Intra- and interpopulation genetic parameter estimates and implications in selection. *Rev. Brasil. Genet.*, 6, pp. 15-27.
- MOREAU, L.; CHARCOSSET, A.; GALLAIS, A. (1998). Experimental evaluation of several cycles of marker-assisted selection in maize. *Euphytica*, 137, pp. 111-118.
- MULAMBA, N.N.; MOCK, J.J. (1978). Improvement of yield potential of the Eto Blanco maize (*Zea mays* L.) population by breeding for plant traits. *Egypt. J. Genet. Cytol.*, 7, pp. 40-51.
- MULLER, H.J. (1923). A simple formula giving the number of individuals required for obtaining one of a given frequency. *Am. Nat.*, 5, pp. 66-73.
- MUIR, W.M. (2007). Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J. Anim. Breed. Genet.*, 124, pp. 342-355.
- NEI, M.; MURATA, M. (1966). Effective population size when fertility is inherited. *Genet. Res.*, 8, pp. 257-260.
- NILSSON-EHLE, N.H. (1909). *Kreuzungsuntersuchungen an hafer und weizen*. Lunds Univversitetis Arskriftt Atd. 2,5, N.o 2: 1-122.
- OTSUKA, Y.; EBERHART, S.A.; RUSSELL, W.A. (1972). Comparisons of prediction for maize hybrids. *Crop Sci.*, 12, pp. 325-331.
- PATERNIANI, E. (1978). Reciprocal recurrent selection based on half-sib progenies and prolific plants in maize (*Zea mays* L.). *Maydica*, 23, pp. 209-219.
- PATERNIANI, E.; VENCOSKY, R. (1977). Reciprocal recurrent selection based on testcross of half sib families. *Maydica*, 22, pp. 141-152.
- PATTERSON, H.D.; THOMPSON, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika*, 58, pp. 545-554.
- PEARSON K. (1904). *On the theory of contingency and its relation to association and normal correlation*. Dulau and Co., Londres. p. 35.
- PEDERSON, D.G. (1981). A least-squares method for choosing the best relative proportions when intercrossing cultivars. *Euphytica*, 30, pp. 153-160.
- PELOQUIN, S.J. (1991). *Chromosomal and cytoplasmic manipulations*. En: *Plant Breeding II*. Editor: Frey, K. Plant Breeding Symposium, 2d; Iowa State University. ISU Press, Ames, Iowa, Estados Unidos de América. p. 497.
- PESEK, J.; BAKER, R.J. (1969). Desired improvement in relation to selection indices. *Can. J. Plant Sci.*, 49, pp. 803-804.
- PIEPHO, H.P.; MÖHRING J.; MELCHINGER, A.E. (2008). BLUP for phenotypic selection in plant breeding. *Euphytica*, 161, pp. 209-228.
- RAWLINGS, J.O.; THOMPSON, D.L. (1962). Performance level as criterion for the choice of maize testers. *Crop Sci.*, 2, pp. 217-220.
- REMINGTON, D.L.; THORNSBERRY, J.M.; MATSUOKAY.; WILSON, L.M.; WHITT, S.R.; DOEBLEY, J.; KRESOVICH, S.; GOODMAN, M.; BUCKLER IV, E.S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *PNAS*, 98, pp. 11479-11484.
- RESENDE, JR., M.F.R.; MUÑOZ, P.; ACOSTA, J.J.; PETER, G.F.; DAVIS, J.M.; GRATTAPAGLIA, D.; RESENDE M.D.V.; KIRST, M. (2011). Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist*. Doi:10.1111/j.1469-8137.2011.03895.x
- ROBERTSON, A. (1960). *A theory on limits of artificial selection*. Proceedings of the Royal Science Society B. DOI:10.1098/rspb.1960.0099

- ROBINSON, H.F.; COMSTOCK, R.E.; HARVEY, P.H. (1949). *Estimates of heritability and the degree of dominance in corn*. Agron. J., 41, pp. 353-359.
- RONNINGEN, K. (1976). *The estimation of appropriate selection intensity from skewed distribution*. Acta Agriculture Scandinavica, 26, pp. 80-84.
- RUSSELL, W.A.; EBERHART, S.A. (1975). *Hybrid performance of selected maize lines from reciprocal recurrent and testcross selection programs*. Crop Sci., 15, pp. 1-4.
- SATTLER, M.; CARVALHO, R., CLARINDO, W. (2016). *The polyploidy and its key role in plant breeding*. Planta 243 (2), pp. 281-296.
- SCHULTZ, JR. E.F. (1955). *Rules of thumb for determining expectations of mean squares in analysis of variance*. Biometrics, 11, pp. 123-135.
- SCHWARZBACH, E. (1972). *Einige anwendungsmöglichkeiten elektronischer datenverarbeitung (EDB) für die beurteilung von zuchtmaterial*. Arb. Tag. Oesterr. Pflanzeng. Gumpenstein, 277-287.
- SEDCOLE, J.R. (1977). *Number of plants necessary to recover a trait*. Crop Sci., 17, pp. 667-668.
- SHUKLA, G.K. (1972). *Some statistical aspects of partitioning genotype-environmental components of variability*. Heredity, 29, pp. 237-245.
- SHULL, G.H. (1908). *The composition of a field of maize*. Am. Breeders' Magazine, 4, pp. 296-301.
- SHULL, G.H. (1909). *A pure-line method in corn breeding*. Am. Breeders' Magazine, 5, pp. 51-59.
- SHULL, G. H. (1910). *Hybridization methods in corn breeding*. Am. Breeders' Magazine, 6, pp. 98-107.
- SHULL, G.H. (1952). *Beginnings of the heterosis concept*. En: Heterosis. Editor: Gowen, J.W. Iowa State University Press. pp. 14-48.
- SMITH, C. (1936). *A discriminant function for plant selection*. Ann. Eugenics, 7, pp. 240-250.
- SMITH, O.S. (1979a). *A model for evaluating progress from recurrent selection*. Crop Sci., 19, pp. 223-226.
- SMITH, O.S. (1979b). *Application of a modified diallel analysis to evaluate recurrent selection for grain yield in maize*. Crop Sci., 19, pp. 819-822.
- SMITH, O.S. (1983). *Evaluation of recurrent selection in BSSS, BSCB1 and BS13 maize populations*. Crop Sci., 23, pp. 35-40.
- SMITH, O.S. (1984). *Comparison of effects of reciprocal recurrent selection in the BSSS (R), BSCB1(R) and BS6 populations*. Maydica, 29, pp. 1-8.
- SMITH, O.S.; HALLAUER, A.R.; RUSSELL, W.A. (1981). *Use of index selection in recurrent selection programs in maize*. Euphytica, 30, pp. 611-618.
- SOLBERG, T.R.; SONESSON, A.K.; WOOLLIAMS, J.A.; MEUWISSEN, T.H.E. (2008). *Genomic selection using different marker types and densities*. Journal of Animal Science, 86, pp. 2447-2454. Doi:10.2527/jas.2007-0010
- SOUZA, JR. C.L. (1987). *Reciprocal recurrent selection with half sib progenies obtained alternatively from non-inbred (S0) and inbred (S1) plants in maize*. Maydica, 32, pp. 19-31.
- SOUZA, JR. C.L. (1992). *Interpopulation genetic variances and hybrid breeding programs*. J. Genetics, 15, Brasil. pp. 643-656.
- SOUZA, JR. C.L. (1993). *Comparisons of intra-, interpopulation, and modified recurrent selection methods*. J. Genetics, 16, Brasil. pp. 91-105.
- SOUZA, JR. C.L. (1999). *Recurrent selection and heterosis*. En: Genetics and exploitation of heterosis in crops. Editores: Coors, J.G.; Pandey, S. ASA-CSSA, Madison, Wisconsin, EE. UU. p. 524.
- SOUZA, JR. C.L.; MIRANDA FILHO, J.B. (1985). *Alteracoes na heterose via selecao intra e interpopulacional*. Pesq. Agropec. Bras., 20, pp. 1197-1201.
- SOUZA, JR. C.J.; ZINSLY, J.R. (1985). *Relative genetic potential of braquitic maize (Zea mays L.) varieties as breeding populations*. Braz. J. Genet., 8, pp. 523-533.
- SOUZA, JR. C.L.; GERALDI, I.O.; VENCOVSKY, R. (2000). *Response to recurrent selection under small effective population size*. Genetics and Molecular Biology, 23, pp. 841-846.
- SPRAGUE, G.F. (1984). *Organization of breeding programs*. Proceedings of the 20th Annu. Illinois Corn Breeders School. Univ. of Illinois, Urbana-Champaign, Illinois, EE. UU. pp. 20-31.
- SPRAGUE, G.G.; MILLER, P.A. (1950). *A suggestion for evaluating current concepts of the genetic mechanism of heterosis in corn*. Agron. J., 42, pp. 161-162.

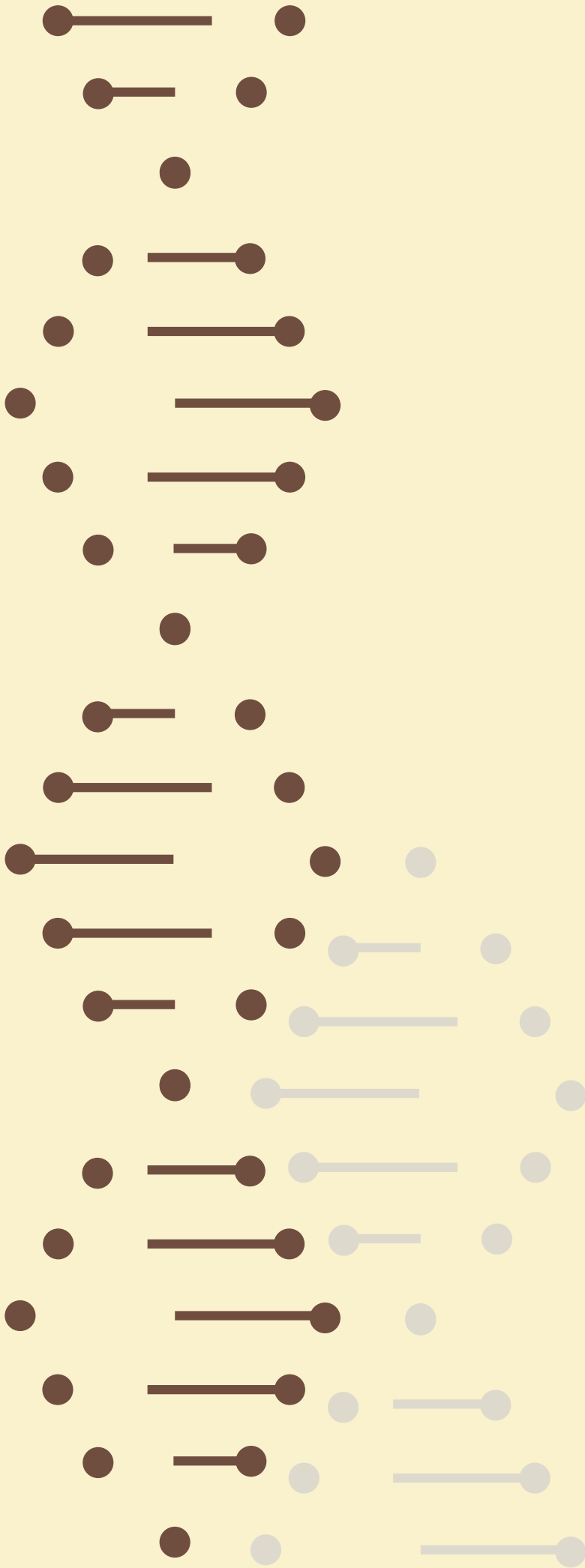
- SPRAGUE, G.; TATUM, L.A. (1942). *General vs. specific combining ability in single crosses of maize*. J. Am. Soc. Agron., 34, pp. 923-932.
- STEBBINS, JR. G.L. (1938). *Cytological characteristics associated with the different growth habits in the dicotyledons*. Amer. J. Bot., 25, pp. 189-198.
- TAI, G.C.C. (1977). *Index selection with desired gains*. Crop Sci., 17, pp. 182-183.
- TALLIS, G.M. (1962). *A selection index for optimum genotype*. Biometrics, 18, pp. 120-122.
- TALLIS, G.M. (1959). *Sampling errors of genetic correlation coefficients calculated from analyses of variance and covariance*. Aust. J. Stat., 1, pp. 35.
- TECHNOW, F.; SCHRAG, T.A.; SCHIPPRACK, W.; BAUER, E.; SIMIANER, H.; MELCHINGER, A. (2014). *Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize*. Genetics, 197, pp. 1343-1355. DOI:10.1534/genetics.114.165860
- VARGAS HERNÁNDEZ, M.; CROSSA, J. (2000). *El análisis AMMI y la gráfica del biplot en SAS*. Unidad de Biometría y Estadística. CIMMYT. México.
- VÉLEZ-TORRES, M.; GARCÍA-ZAVALA, J.J.; HERNÁNDEZ-RODRÍGUEZ, M.; LOBATO-ORTIZ, R.; LÓPEZ-REYNOSO, J.J.; BENÍTEZ-RIQUELME, I.; MEJÍA-CONTRERAS, J.A.; ESQUIVEL-ESQUIVEL, G.; MOLINA-GALÁN, J.D.; PÉREZ-RODRÍGUEZ, P.; ZHANG, XUECAI. (2018). *Genomic prediction of the general combining ability of maize lines (Zea mays L.) and the performance of their single crosses*. Plant Breeding, 137, pp. 379-387. DOI:10.1111/pbr.12597
- VENCOVSKY, R. (1978). *Effective size of monoecious populations submitted to artificial selection*. Rev. Brasil. Genet., 1, pp. 181-191.
- VENCOVSKY, R. (1980). *Herencia Cuantitativa*. En: Melhoramento e producao do milho no Brasil. Editor: Paterniani, E. ESALQ - Fundacao Cargill, Piracicaba, Brasil. p. 650.
- VENCOVSKY, R.; BARRIGA, P. (1992). *Genética biométrica no fitomelhoramento*. Riverirao Preto. Sociedade Brasileira de Genética. p. 496.
- WARDYND, B.M.; EDWARDS, J.W.; LAMKEY, K.R. (2007). *The genetic structure of a maize population: the role of dominance*. Crop. Sci., 47, pp. 467-474.
- WALSH, B. (2002). *Quantitative genetics, genomics, and the future of plant breeding. Quantitative genetics, genomics and plant breeding*. CABI Publishing. p. 400. (Disponible: <https://pdfs.semanticscholar.org/9d69/68c92ee620743381a7f9739bcb216b81d6dc.pdf> verificado: 8 de agosto de 2020).
- WALSH, B. (2012). *Mixed-model estimation of genetic variances*. Lecture notes, Uppsala EQG 2012 course. Versión 28 de enero de 2012.
- WATTERSON, G.A. (1962). *Some theoretical aspects of diffusion theory in population genetics*. Ann. Math. Stat., 33, pp. 939-957.
- WEINBERG, W. (1908). *Über den Nachweis der Vererbung beim Menschen*. Jahreshfte des vereins für vaterländische Naturkunde in Württemberg, 64, pp. 368-382.
- WILLIAMS, J.S. (1962). *The evaluation of a selection index*. Biometrics, 18, pp. 375-393.
- WHITTAKER, J.C. (2001). *Marker assisted selection and introgression*. En: Handbook of Statistical Genetics. Editores: Balding, D.J.; Bishop, M.; Cannings, C. John Wiley, New York, EE. UU. pp. 673-695.
- WRICKE, G.; WEBER W.E. (1986). *Quantitative genetics and selection in plant breeding*. Walter de Gruyter. Berlín, Nueva York. p. 406.
- WRIGHT S. (1921). *Systems of mating. II. The effects of inbreeding on the composition of a population*. Genetics, 6, pp. 124-143.
- WRIGHT S. (1922). *Coefficients of inbreeding and relationship*. The American Naturalist, 56, pp. 330-338.
- WRIGHT S. (1923). *The theory of path coefficients: A reply to Niles' criticism*. Genetics, 8, pp. 239-255.
- WRIGHT S. (1931). *Evolution in Mendelian populations*. Genetics, 16, pp. 97-159.
- WRIGHT, S.; NESS, R.W.; FOXE, J.P.; BARRETT, S.C.H. (2008). *Genomic consequences of outcrossing and selfing in plants*. Int. J. Plant Sci., 169, pp. 105-118.
- XAVIER, A.; MUIR, W.M.; CRAIG, B.; RAINEY, K.M. (2016). *Walking through the statistical black boxes of plant breeding*. Theor. Appl. Genet., 129, pp. 1933-1949.
- YAN, W. Y HUNT, L.A. (2002). *Biplot analyses of multi-environment trial data*. En: Quantitative genetics, genomics, and plant breeding. Ed.: Kang, M.S. CABI Publishing, Nueva York, EE. UU. p. 400.



YOUNG, S.S. Y. (1961). *A further examination of the relative efficiency of three methods of selection for genetic gains under less restricted conditions*. Genet. Res., 2, pp. 106-121.

YULE, G.U. (1906). *On the theory of inheritance of quantitative compound characters on the basis of Mendel's laws*. Third Int. Conf. Genet, pp. 140-142.

ZANONI, R.; DUDLEY, J.W. (1989). *Use of F2 generations to identify inbreds with favorable alleles not present in an elite hybrid*. Crop Sci., 29, pp. 583-589.



A.1. Medidas de tendencia central

A.1.1. Media aritmética o promedio aritmético

μ = media poblacional o paramétrica

\bar{y} = media muestral

N = número de individuos (observaciones) en la población

n = número de individuos (observaciones) en la muestra

$$\mu = \frac{\sum_{i=1}^N y_i}{N},$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

Otra manera de calcular la media es calcular la media ponderada, es decir, sumando el producto de los valores observados (y_i) por el número de veces que cada valor ocurre en la muestra (f_i).

$$\mu = (f_1 y_1 + f_2 y_2 + f_3 y_3 + \dots + f_k y_k) / (f_1 + f_2 + f_3 + \dots + f_k)$$

Si expresamos $f_1, f_2, f_3, \dots, f_k$ en términos de frecuencias:

$$f_1 / (f_1 + f_2 + f_3 + \dots + f_k) = f_1 / N = p_1$$

$$f_2 / (f_1 + f_2 + f_3 + \dots + f_k) = f_2 / N = p_2$$

$$f_3 / (f_1 + f_2 + f_3 + \dots + f_k) = f_3 / N = p_3$$

$$f_k / (f_1 + f_2 + f_3 + \dots + f_k) = f_k / N = p_k$$

Entonces:

$$\mu = (p_1 y_1 + p_2 y_2 + p_3 y_3 + \dots + p_k y_k)$$

Los desvíos entre los valores individuales y la media muestral, o desvío dentro de la muestra es:

$$(y - \bar{y})$$

Por propiedad de la media aritmética,

$$\sum_{i=1}^n (y - \bar{y}) = 0$$

A.1.2. Otras medidas de tendencia central

Cuando la distribución de las observaciones está sesgada hacia la derecha o hacia la izquierda resulta más apropiado utilizar la mediana o la moda para reflejar las medidas de posición o tendencia central de una población o su muestra. La mediana es el valor que deja el 50 % de las observaciones a uno y otro lado de la distribución. La moda es el valor más frecuente en la distribución (Figura A.1).

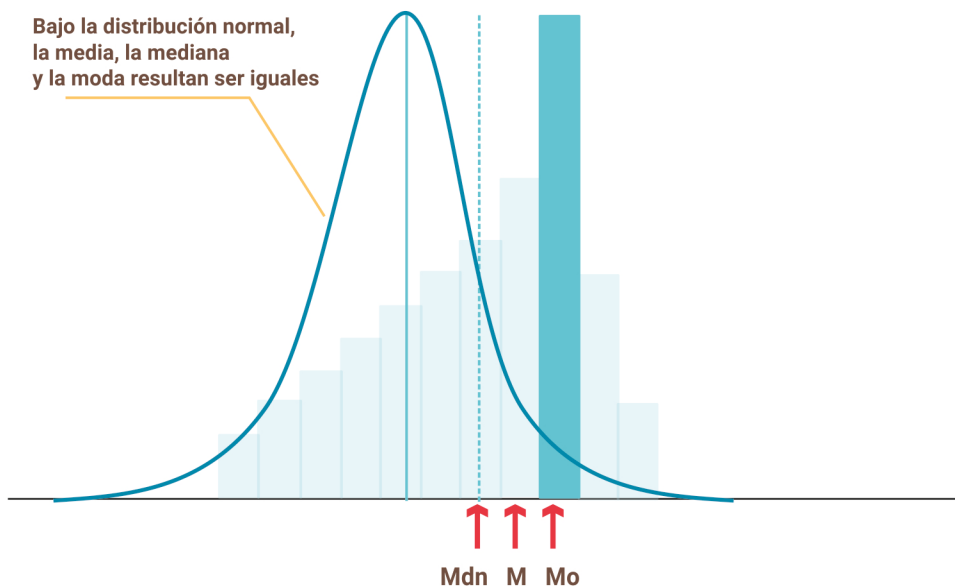


Figura A.1

Ubicación relativa de medidas de tendencia central [media (M), mediana (Mdn) y moda (Mo)] bajo dos condiciones de curtosis.

A.1.2.1. Media geométrica (G)

$$G = (y_1 y_2 y_3 \dots y_N)^{1/N}$$

donde N es el número de observaciones.

A.1.2.2. Media armónica (H)

$$1/H = 1/N \sum 1/y_i$$

A.2. Medidas de dispersión

A.2.1. Variancia

La variancia es la suma de cuadrados de los desvíos entre las observaciones y la media. Esta suma de cuadrados se divide por los grados de libertad.



σ^2 = variancia poblacional o paramétrica; σ = desvío estándar poblacional

S_2 = variancia muestral; S = desvío estándar muestral

$$\sigma^2 = \sum (y_i - \mu)^2 / N \text{ (en poblaciones de tamaño finito)}$$

$$\sigma^2 = \sum (y_i - \mu)^2 \text{ (en poblaciones cuando se trabaja con frecuencias relativas)}$$

$$S^2 = \sum (y_i - \bar{y})^2 / n - 1 \text{ (en muestras de tamaño } n)$$

La variancia poblacional de una distribución de medias muestrales de tamaño n será:

$${}_n\sigma^2 = \sigma^2 / n \text{ La variancia de una distribución de medias muestrales de tamaño } n \text{ será:}$$

$${}_nS^2 = S^2 / n$$

Puede observarse que las medias muestrales tienen menor grado de dispersión (variancia) que las observaciones simples. Los valores observados bajo el supuesto de distribución normal, para poblaciones con igual media, se ubican a la misma distancia medida en desvíos estándares de esta, cualquiera sean las variancias de las poblaciones (Figuras A.2). Para un mismo valor medido en desvíos estándares, los valores observados serán menores para la población con menor variancia. La probabilidad de ocurrencia de valores observados superiores al indicado será igual para cualquiera de las poblaciones ($P \geq X = P \geq Y = P \geq Z$) (Figura A.2).

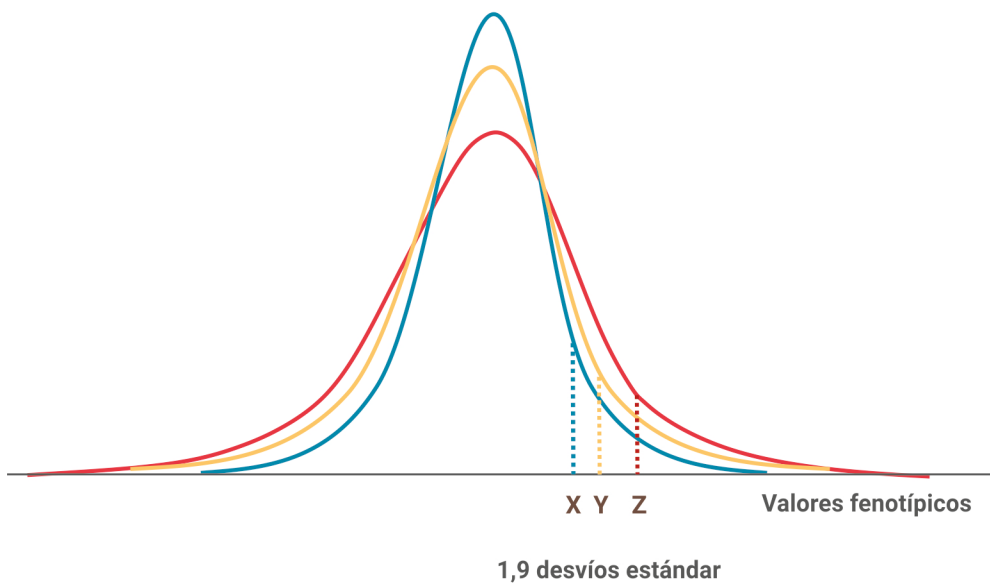


Figura A.2. Ubicación de valores observados medidos en unidades estándares correspondientes a tres poblaciones con igual media y distinta variancia.

A.2.2. Coeficiente de variación

El coeficiente de variación muestral expresado en porcentaje es:

$$CV\% = 100 S / \bar{y}$$

El coeficiente de variación poblacional es

$$CV\% = 100 \sigma/\mu$$

Con frecuencia se interpreta al coeficiente de variación muestral como un índice de confiabilidad de datos experimentales. Ello no es necesariamente así en todos los casos, puesto que el coeficiente de variación está inversamente correlacionado con la media muestral.

A.3. Medidas de asociación entre variables

La covariancia, la regresión y la correlación proveen una medida del nivel de asociación entre variables aleatorias.

La covariancia es la suma de los productos medios entre las observaciones de las variables X e Y:

$$\text{Cov}(x,y) = x_1y_1 + x_2y_2 + x_3y_3 + \dots + x_ny_n - [\sum x \sum y/N]$$

$$\text{Cov}(x,y) = x_1y_1 + x_2y_2 + x_3y_3 + \dots + x_ny_n - [\sum x \sum y/n-1]$$

La covariancia entre dos variables aleatorias depende de las unidades en las que ambas se miden. Una manera de comparar el grado de asociación entre pares de variables que resulte libre del efecto de las unidades de medida o de diferencias de escala es calcular la covariancia entre variables estandarizadas. De tal manera:

$$\begin{aligned} \text{Cov}(x,y) = & [(x_1-X)/S_x][(y_1 - Y)/S_y] + [(x_2-X)/S_x][(y_2 - Y)/S_y] + [(x_3-X)/S_x][(y_3 - Y)/S_y] + \\ & \dots + [(x_n-X)/S_x][(y_n - Y)/S_y] \end{aligned}$$

Donde $(x_i - X)$ son los desvíos de los valores observados de x respecto de su media, S_x es el desvío estándar de la variable x, e $(y_i - Y)$ son los desvíos de los valores observados de y_i respecto de su media, y S_y es el desvío estándar de la variable y.

En consecuencia la variancia de las variables estandarizadas es igual a 1. La covariancia entre las variables estandarizadas tendrá un valor acotado en el rango -1 a +1 y se lo denomina coeficiente de correlación.

A.3.1. Regresión

La regresión entre una variable independiente x y una variable dependiente y mide el cambio esperado en y por unidad de cambio en x:

$$b_{yx} = \text{Cov}(y,x)/S_x^2$$

El modelo de regresión lineal simple expresado en notación matricial es

$$Y = X \beta + \epsilon ,$$

donde Y es el vector de observaciones de la variable dependiente, X es la matriz de observaciones de la variable independiente y la columna de incidencia de la media general, β el vector de parámetros de regresión y ϵ es el vector del error experimental asociado a cada una de las observaciones.

Es decir:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \dots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \dots & \dots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \dots \\ e_n \end{pmatrix}$$

El vector de estimadores de parámetros (b) se calcula despejando dicho término del sistema de ecuaciones normales:

$$(X'X) * b = X'Y$$

$$b = (X'X)^{-1} * X'Y$$

A.3.2. Correlación

Mide el grado de asociación entre dos variables aleatorias x e y mediante:

$$r_{xy} = \text{Cov}(x,y) / S_x S_y$$

Recordar que el coeficiente de correlación equivale a la covariancia entre dos variables estandarizadas, por lo que mide el grado de asociación entre variables libre de los efectos de diferentes escalas de medida.

El coeficiente de determinación será:

$$R_{xy} = r_{xy}^2$$

A.3.3. Relación entre regresión y correlación simple

Dados

$$b_{yx} = \text{Cov}(y,x) / S_x^2$$

$$b_{xy} = \text{Cov}(x,y) / S_y^2$$

$$b_{yx} b_{xy} = [\text{Cov}(y,x)]^2 / S_x^2 S_y^2 = R_{xy} = r_{xy}^2$$

$$(b_{yx} b_{xy})^{1/2} = r_{xy}$$

A.3.4. Regresión lineal múltiple

En los modelos de regresión lineal múltiple, una variable aleatoria dependiente es función de dos o más variables aleatorias independientes:

$$y = b_0 x_0 + b_1 x_1 + b_2 x_2 + e$$

El modelo expresado en términos matriciales es semejante al de la regresión lineal simple, pero la dimensión de la matriz X es mayor (el número de columnas es igual al número de variables independientes más 1) y el vector de parámetros tiene más filas, equivalentes al número de columnas de la matriz X:

$$(X'X) \cdot b = X'Y$$

$$b = (X'X)^{-1} \cdot X'Y$$

La matriz X'X será:

$$\begin{pmatrix} \sum x_0^2 & \sum x_0x_1 & \sum x_0x_2 \\ \sum x_1x_0 & \sum x_1^2 & \sum x_1x_2 \\ \sum x_2x_0 & \sum x_2x_1 & \sum x_2^2 \end{pmatrix}$$

Y la matriz X'Y será:

$$\begin{pmatrix} \sum x_0y \\ \sum x_1y \\ \sum x_2y \end{pmatrix}$$

La suma de cuadrados del modelo completo que considera k variables aleatorias de las que dependen los valores observados Y_i será:

$$SC(\text{modelo completo}) = SC(b_0, b_1 \dots b_k) = b' \cdot X'Y$$

La suma de cuadrados de la regresión múltiple debida a k parámetros será:

$$SC(\text{regresión múltiple}) = SC(b_1, \dots, b_k) = b' \cdot X'Y - (\sum Y_i)^2/n$$

El último término del miembro de la derecha de la igualdad representa la suma de cuadrados del modelo reducido, que asume que $Y = \beta_0 + \epsilon$, donde b_0 es $(\sum Y_i)/n$, es decir, el promedio de las observaciones de la variable dependiente.

Notar que en el modelo completo, b_0 es diferente:

$$b_0 = (\sum Y_i)/n + b_1(\sum X_{i1})/n + b_2(\sum X_{i2})/n$$

La suma de cuadrados de los residuales será:

$$SC(\text{residuales}) = Y'Y - b' \cdot X'Y$$

La variancia del vector b de estimadores de los parámetros es:

$$\text{Var}(b) = (X'X)^{-1} \cdot s^2,$$

donde s^2 es la variancia del error experimental.



A.3.5. Correlación parcial

Dada una serie de variables, la correlación parcial entre dos de ellas es la correlación entre ambas, manteniendo constantes o fijas al resto. Dadas las variables aleatorias $x_1, x_2, x_3, \dots, x_p$, la correlación parcial, por ejemplo $r_{14.23}$, es la correlación entre x_1 y x_4 ajustadas por x_2 y x_3 .

La matriz (R) de correlaciones entre variables es:

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1p} \\ r_{21} & 1 & r_{23} & \dots & r_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & r_{p3} & \dots & 1 \end{pmatrix}$$

y su matriz inversa:

$$C = \begin{pmatrix} c_{11} & c_{12} & c_{13} & \dots & c_{1p} \\ c_{21} & c_{22} & c_{23} & \dots & c_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ c_{p1} & c_{p2} & c_{p3} & \dots & c_{pp} \end{pmatrix}$$

La correlación parcial entre dos variables x_i y x_j (manteniendo fijas a las restantes variables) es:

$$r_{ij..} = -c_{ij} / (c_{ii} c_{jj})^{0.5}$$

La prueba de significancia de los coeficientes de correlación parcial ($H_0: \rho_{ij..} = 0; H_1: \rho_{ij..} \neq 0$) sigue una distribución t, cuyo valor calculado es:

$$t_{(n-p)} = r_{ij..} (n-p)^{0.5} / (1 - r_{ij..}^2)^{0.5}$$

donde n es el número de observaciones, p el número de variables aleatorias y (n-p) son los grados de libertad.

A.3.6. Correlación múltiple

El coeficiente de correlación múltiple mide el ajuste entre los valores observados para una variable dependiente y los valores estimados de esa variable en función de un modelo de regresión múltiple respecto de un número de variables independientes. El coeficiente de correlación múltiple es la raíz cuadrada del coeficiente de determinación.

A.3.7. Regresión utilizando variables estandarizadas

Cuando las observaciones de las variables se expresan como desvíos respecto a su media aritmética y estos expresados en relación con el desvío estándar, cada variable estandarizada tendrá una media igual a 0 y una variancia y desvío estándar iguales a 1.

En el caso de una regresión lineal múltiple con tres variables aleatorias, el modelo es:

$$y = b_0 x_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + e$$

La variable x_0 corresponde a la ordenada al origen y su valor es 1 para cada una de las observaciones. Cuando se procede a estandarizar todas las variables, el modelo es:

$$(y-Y)/S_y = (x_1 - X_1)/S_{x1} + (x_2 - X_2)/S_{x2} + (x_3 - X_3)/S_{x3}$$

Reexpresando a cada término estandarizado respectivamente como y' , x_1' , x_2' y x_3' , el modelo es:

$$y' = b_1'x_1' + b_2'x_2' + b_3'x_3'$$

donde:

$$b_1' = b_1 (S_{x1}/S_y)$$

$$b_2' = b_2 (S_{x2}/S_y)$$

$$b_3' = b_3 (S_{x3}/S_y)$$

La estandarización simplifica algunas expresiones (Cox, 1986). Así por ejemplo, la sumatoria de los desvíos respecto a la media elevados al cuadrado, es igual simplemente a la suma de los valores estandarizados al cuadrado [$\sum(y' - Y')^2 = \sum(y')^2$] y como la variancia de todas las variables es igual a 1, para cada par de variables, covariancia, la correlación y los coeficientes de regresión son iguales entre sí.

El modelo de regresión de variables estandarizadas expresado en notación matricial se simplifica de la siguiente manera:

$$(X'X) b' = X'Y,$$

donde la matriz $X'X$ es

$$\begin{pmatrix} \sum x_1^2 & \sum x_1x_2 & \sum x_1x_3 \\ \sum x_2x_1 & \sum x_2^2 & \sum x_2x_3 \\ \sum x_3x_1 & \sum x_3x_2 & \sum x_3^2 \end{pmatrix}$$

Y la matriz $X'Y$:

$$\begin{pmatrix} \sum x_1y \\ \sum x_2y \\ \sum x_3y \end{pmatrix}$$

El vector de estimadores de los parámetros de regresión es:

$$\begin{pmatrix} b_1' \\ b_2' \\ b_3' \end{pmatrix}$$

y puede reexpresarse este sistema de la siguiente manera:

$$\begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix} \begin{pmatrix} b_1' \\ b_2' \\ b_3' \end{pmatrix} = \begin{pmatrix} r_{1y} \\ r_{2y} \\ r_{3y} \end{pmatrix}$$

Por lo tanto el sistema de ecuaciones es:

$$b_1 + b_2 r_{12} + b_3 r_{13} = r_{1y}$$

$$b_1 r_{21} + b_2 + b_3 r_{23} = r_{2y}$$

$$b_1 r_{31} + b_2 r_{32} + b_3 = r_{3y}$$

A.4. Modelos lineales

Los modelos son versiones simplificadas de una realidad o fenómeno complejo que ayudan a explicar y a predecir tal fenómeno a partir de sus variables más importantes y sus interrelaciones. Los modelos estocásticos son aquellos que poseen al menos una variable aleatoria. El modelo lineal general (MLG o en inglés, GLM) se puede representar como:

$$Y = f_{(X)} + \varepsilon,$$

donde ε es una variable aleatoria (por lo tanto Y también lo es).

Y es función de al menos otra variable X no aleatoria. La función de X puede resultar de una combinación de parámetros y efectos fijos de X (o niveles de X). En la medida que los parámetros de la función de X puedan ser estimados, será posible estimar Y. En el caso de poder estimar Y, (Y_e), es posible estimar los residuales que responden a la variable aleatoria ε partir de:

$$e = Y_e - f_{(X)}$$

La función de X puede tomar diferentes formas, como por ejemplo una regresión lineal simple o múltiple de la forma:

$$f_{(X)} = \beta_0 + \beta_1 X_1, \text{ o bien}$$

$$f_{(X)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2,$$

o de un modelo de diseño experimental tal como:

$$f_{(X)} = b_i + t_j,$$

donde b_i es el efecto de la i-ésima repetición o bloque completo y t_j es el efecto de la j-ésima variedad evaluada en un experimento con i repeticiones.

La condición de linealidad del modelo está dada por cuanto la variable aleatoria Y depende de una función lineal de los parámetros ($\beta_0, \beta_1, \beta_2, b_i, t_j$).

Supongamos el MLG expresado en forma matricial del tipo:

$$Y = X\beta + \varepsilon,$$

donde n es el número de observaciones disponibles,

Y es el vector de valores observados, de dimensión $(n \times 1)$,

β el vector de parámetros para estimar, de dimensión $(p \times 1)$,

X es la matriz de incidencia o de diseño, de dimensión $(n \times p)$, y

ϵ es el vector de la variable aleatoria del error, de dimensión $(n \times 1)$.

El modelo es de rango completo siempre que $n \geq p$. Se asume que los errores de las observaciones son independientes entre sí y que tienen una variancia común σ_e^2 .

En el modelo de rango completo la matrix $X'X$ es no-singular y existe una solución única para el vector de parámetros β .

En el modelo de rango incompleto, el rango de la matrix X es $r < p$, la matrix $X'X$ es singular y por lo tanto existen infinitas soluciones para el vector de parámetros.

A.4.1. Medias de mínimos cuadrados (LSMeans)

Cuando se dispone de datos desbalanceados, las medias de mínimos cuadrados son estimaciones de las medias aritméticas que podrían esperarse obtener si las clases o subclases hubieran tenido el mismo número de observaciones.

Asumiendo un modelo con dos efectos, E (ambientes) y G (genotipos), con tres niveles y dos niveles respectivamente y los datos presentados en la Tabla A.1.

Genotipos	Ambiente			N	Media	Media	
	E1	E2	E3		aritmética	mínimos cuadrados	
G1	20, 40, 30	30, 40	35	6	32,5	-9	31
G2	60, 50	55, 55	-	4	55,5	13,5	53,5
N	5	4	1	10	43,75	2,25	
Media aritmética	40	45	35	40	41,5		42,25
Media mínimos cuadrados	42,25	47,25	37,25		42,25		

Tabla A.1.

Comparación de medias aritméticas y medias de mínimos cuadrados aplicada a la evaluación de dos genotipos en tres ambientes.



Las medias de mínimos cuadrados de G1 y de G2 se calculan como:

$$\text{LSMean } G_i = \mu + g_i + e.$$

donde g_i es el efecto del genotipo i -ésimo y e es el efecto promedio de los ambientes a través de los dos genotipos:

$$\text{LSMean } G_1 = 41,5 + (-9,0) + (-1,5) = 31,0$$

$$\text{LSMean } G_2 = 41,5 + (13,5) + (-1,5) = 53,5$$

La media de mínimos cuadrados de E1, E2 y E3 se calculan como:

$$\text{LSMean } E_j = \mu + e_j + g.$$

donde e_j es el efecto del ambiente j -ésimo y g es el efecto promedio de los genotipos a través de los tres ambientes:

$$\text{LSMean } E_1 = 41,5 + (-1,5) + (2,25) = 42,25$$

$$\text{LSMean } E_2 = 41,5 + (3,5) + (2,25) = 47,25$$

$$\text{LSMean } E_3 = 41,5 + (-6,5) + (2,25) = 37,25$$

Nótese que las medias de mínimos cuadrados son similares, pero distintas a las respectivas medias aritméticas. Si los datos hubieran sido balanceados, las medias aritméticas y las de mínimos cuadrados coincidirían.

A.4.2. Variancia de funciones lineales

La variancia de una función lineal Y de variables aleatorias puede calcularse siguiendo las siguientes reglas:

1. Sustituir la variable Y por su expresión lineal.
2. Si existe algún término que es una constante que no aparece multiplicando o dividiendo a una variable aleatoria, eliminarlo.
3. Realizar el cuadrado de la función remanente que describe a Y .
4. Reemplazar los términos resultantes por variancias si son potencias al cuadrado, o por covariancias si son productos cruzados.

Ejemplos:

a. Dada la función lineal del tipo

$$Y = a + b$$

donde a y b son variables aleatorias:

$$V(Y) = V(a + b)$$

Se eleva al cuadrado la función $(a + b)$:

$$(a + b)^2 = a^2 + b^2 + 2ab$$

$$V(Y) = V(a + b) = V(a) + V(b) + 2COV(a,b)$$



b. Dada la función del tipo:

$$Y = a + bx$$

donde a y b son constantes y x es una variable aleatoria:

$$V(Y) = V(a + bx)$$

Se elimina la constante a y se eleva al cuadrado la expresión remanente:

$$(bx)^2 = b^2x^2$$

$$V(Y) = b^2V(x)$$

c. Dada la función del tipo:

$$Y = a + bx + cz$$

donde a, b y c son constantes, x y z variables aleatorias.

$$V(Y) = V(a + bx + cz)$$

Se eleva al cuadrado el término (bx + cz):

$$(bx + cz)^2 = b^2x^2 + c^2z^2 + 2bcxz$$

$$V(Y) = b^2 V(x) + c^2 V(z) + 2bcCOV(x,z)$$

A.4.3. Covariancia de funciones lineales

La covariancia entre dos funciones lineales, X e Y, puede calcularse siguiendo las siguientes reglas:

1. Sustituir cada función, X e Y, por su expresión lineal.
2. Si existe algún término que es una constante que no aparece multiplicando o dividiendo a una variable aleatoria, eliminarlo.
3. Realizar el producto cruzado entre las expresiones remanentes de X e Y.
4. Reemplazar cada término resultante que sea producto de las dos variables aleatorias por covariancias.

Ejemplo:

Dados $X = (a + b)$ e $Y = (c + d)$, siendo a, b, c y d variables aleatorias:

$$COV(XY) = COV(a + b, c + d)$$

$$(a + b)(c + d) = ac + ad + bc + bd$$

$$COV(X,Y) = COV(a,c) + COV(a,d) + COV(b,c) + COV(b,d)$$

A.5. Estructura de tratamientos

La estructura de tratamientos que se evalúan en experimentos puede ser de tipo anidado o de tipo cruzado. La característica de la estructura anidada es que cada tratamiento queda definido por al menos dos factores clasificatorios y los niveles de uno de ellos es diferente para cada uno de los niveles del otro factor. Por el contrario, la estructura de tratamientos es de tipo cruzado cuando los niveles de cada uno de los factores clasificatorios se aplican a todos los niveles de los demás factores clasi-

ficatorios (Figura A.3). En la estructura anidada no es posible determinar la interacción entre los dos factores clasificatorios que definen cada tratamiento, mientras que en los diseños con estructura de tratamientos de tipo cruzado es posible determinar las interacciones entre ellos.

[a]	Factor P			[b]	Factor P		
	P1	P2	P3		P1	P2	P3
Factor Q	Q1	Q3	Q5	Factor Q	Q1	Q1	Q1
	Q2	Q4	Q6		Q2	Q2	Q2

Figura A.3. Esquema de diseño de tratamientos anidado [A] y cruzado [B] cuando los tratamientos están definidos por dos factores, P y Q.

A.6. Esperanzas de los cuadrados medios en el análisis de variancia

En cualquier análisis de variancia es preciso conocer la esperanza de los cuadrados medios de las fuentes de variación para poder determinar cuál es el denominador de las pruebas F de significancia que debe emplearse. La esperanza de un cuadrado medio para una fuente de variación es una función lineal de las variancias o sumatorias de efectos (ponderadas por unas constantes que dependen de las características del experimento) que son explicatorias de ese cuadrado medio. En los diseños genéticos, como veremos más adelante, es importante conocer la esperanza de los cuadrados medios para poder estimar componentes de variancia fenotípica y genotípica.

Las esperanzas de los cuadrados medios de cada fuente de variación pueden ser diferentes en función del modelo y de la estructura de los datos experimentales que se asuma:

- Modelo I o de efectos fijos: todos los factores de variación incluidos en el modelo tienen efectos fijos.
- Modelo II o de efectos aleatorios: todos los factores de variación incluidos en el modelo son variables aleatorias, excepto la media general.
- Modelo III o modelo mixto: incluye factores aleatorios y factores fijos.

Cuando decimos que un factor es aleatorio, queremos significar que sigue una distribución de probabilidad. La estructura del experimento se refiere a las variables clasificatorias (repeticiones, ambientes, tratamientos, niveles de tratamiento, etc) y puede ser de tipo cruzada o anidada.

A.6.1. Reglas de Schultz

Para determinar la esperanza de los cuadrados medios, Schultz (1955) elaboró una serie de reglas (*thumb rules*). Cuando todos los efectos del modelo son aleatorios:

Regla 1

El primer paso es listar todas las fuentes de variación en el análisis de variancia y determinar cuáles variables clasificatorias son fijas y cuáles son aleatorias. Normalmente las repeticiones y ambientes son considerados efectos aleatorios y los tratamientos (genotipos) pueden ser considerados aleatorios o fijos. Serán fijos cuando el fitomejorador solo está interesado en conocer los efectos de los g genotipos. Nuestro universo de interés en los genotipos se reduce pura y exclusivamente a aquellos incluidos en el experimento. Serán aleatorios si se considera que esos genotipos no son más que una muestra aleatoria de una población de genotipos más amplia.

Si una fuente de variación corresponde a una interacción entre dos variables clasificatorias, tal interacción deberá considerarse aleatoria toda vez que al menos una de esas variables clasificatorias sea aleatoria. La interacción entre dos o más variables clasificatorias de efectos fijos, también será fija.

Regla 2

2.1. Para cada fuente de variación se debe incluir en la esperanza de su cuadrado medio al componente respectivo y la variancia del error.

2.2. El símbolo del componente correspondiente debe estar identificado por una serie de subscriptos que incluye todos los factores necesarios para describirlo inequívocamente. Los demás subscriptos estarán acompañando al componente correspondiente como constantes. Deben incluirse como componente de la esperanza de los cuadrados medios a todos aquellos cuyos subscriptos identificatorios están incorporados como subscriptos en el componente correspondiente.

Cuando en el modelo hay tanto efectos aleatorios como fijos, hay que aplicar una regla adicional:

Regla 3

Luego de aplicar las reglas 1 y 2 hay que determinar cuáles términos de la esperanza del cuadrado medio deben mantenerse como tales y cuáles deben eliminarse. Para ello el procedimiento es el siguiente:

3.1. Analizar los subscriptos de cada término componente, excepto el específico del componente y determinar si uno o más de ellos corresponden a factores fijos. En tal caso, ese término componente debe eliminarse de la esperanza del cuadrado medio.

3.2. Analizar todos los subscriptos de cada término componente, incluyendo el específico del componente y determinar si todos ellos corresponden a factores fijos, ese término debe ser reemplazado por el promedio de los efectos al cuadrado.

Ejemplos

Evaluación de g genotipos en un experimento con b bloques completos aleatorizados.

Aplicación de regla 1:

Fuentes de variación:

Genotipos $\rightarrow G$ (aleatorio)

Bloques $\rightarrow B$ (aleatorio)

Aplicación de regla 2

Genotipos: 1, 2 ... g

Bloques: 1, 2 ... b

Parcelas: $b_{11}, b_{12}, b_{13} \dots b_{1g}; b_{21}, b_{22}, b_{23} \dots b_{2g}$



Fuente de Variación	Regla 2.1	Regla 2.2	Regla 2.2
Bloques	σ_{GB}^2	$\sigma_e^2 + g\sigma_{GB}^2$	$\sigma_e^2 + g\sigma_B^2$
Genotipos	σ_{BG}^2	$\sigma_e^2 + b\sigma_{GB}^2$	$\sigma_e^2 + b\sigma_G^2$
Error Experimental	σ_e^2		

Evaluación de g genotipos en un experimento con b bloques completos aleatorizados.

Aplicación de regla 1:

Fuentes de variación:

Genotipos → G (fijo)

Bloques → B (aleatorio)

Aplicación de regla 2

Genotipos: 1, 2 ... g

Bloques: 1, 2 ... b

Parcelas: $b_{11}, b_{12}, b_{13} \dots b_{1g}; b_{21}, b_{22}, b_{23} \dots b_{2g}$

Fuente de Variación	Regla 2.1	Regla 2.2	Regla 2.2	Regla 3
Bloques	σ_B^2	$\sigma_e^2 + g\sigma_{GB}^2$	$\sigma_e^2 + g\sigma_B^2$	$\sigma_e^2 + g\sigma_B^2$
Genotipos	σ_G^2	$\sigma_e^2 + b\sigma_{GB}^2$	$\sigma_e^2 + b\sigma_G^2$	$\sigma_e^2 + b\sigma_G^2 + b\sum G_i^2/g-1$
Error Experimental	σ_e^2	σ_e^2	σ_e^2	σ_e^2

Evaluación de g genotipos en un experimento con b bloques completos aleatorizados, en a ambientes.

Aplicación de regla 1:

Fuentes de variación:

Ambientes → A (aleatorio)

Bloques → B (aleatorio)

Genotipos → G (aleatorio)

Ambientes x Genotipos → AxG (aleatorio)

Error Experimental → E Aleatorio

Aplicación de regla 2

Subscriptos				
Ambiente	A	σ_A^2	σ_A^2	
Bloque	B(A)	$\sigma_{B(A)}^2$	$\sigma_{B(A)}^2$	
Genotipo	G	σ_{ABG}^2	$ab\sigma_G^2$	$\sigma_E^2 + b\sigma_{AG}^2 + ab\sigma_G^2$
Amb. x Gen.	ABG	σ_{ABG}^2	$b\sigma_{ABG}^2$	$\sigma_E^2 + b\sigma_{AG}^2$
Error		σ_E^2	σ_E^2	σ_E^2

Evaluación de g genotipos en un experimento con b bloques completos aleatorizados, en a ambientes.

Aplicación de regla 1:

Fuentes de variación:

Ambientes → A (aleatorio)

Bloques → B (aleatorio)

Genotipos → G (fijo)

Ambientes x Genotipos → AxG (aleatorio)

Error Experimental → E Aleatorio

Fuente de Variación	Subscriptos	Aplicando regla 2		Aplicando regla 3
Ambiente	A	σ_A^2	σ_A^2	
Bloque	B(A)	$\sigma_{B(A)}^2$	$\sigma_{B(A)}^2$	
Genotipo	G	σ_{ABG}^2	$ab\sigma_G^2$	$\sigma_E^2 + b\sigma_{AG}^2 + ab\sum G_i^2 / (g-1)$
Amb. x Gen.	ABG	σ_{ABG}^2	$b\sigma_{ABG}^2$	$\sigma_E^2 + b\sigma_{AG}^2$
Error		σ_E^2	σ_E^2	σ_E^2



A.6.2. Modelo de efectos fijos

El modelo de efectos fijos toma la expresión:

$$y = Xb + e$$

donde X es una matriz de diseño o incidencia, b es un vector de efectos fijos contemplados en el modelo y que se pretende estimar y e es el vector de la variable aleatoria del error. Es decir que las únicas variables aleatorias en el modelo son y y e.

La variancia del error (V_e) coincide con la variancia de y (V_y):

$$V_e = V_y = I\sigma_e^2$$

Mientras que la variancia del estimador de y ($V(*y)$) es:

$$V(*y) = P(P'P)^{-1}P'\sigma_e^2$$

La variancia del error (V_e) admite diferentes situaciones:

$$V_e = \begin{bmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_e^2 & 0 \\ 0 & 0 & \sigma_e^2 \end{bmatrix}$$

$$V_e = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}$$

$$V_e = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$$

En el primer caso los errores son independientes e iguales para cada efecto, en el segundo los errores son independientes y específicos para cada efecto y en el tercero los errores de distintos efectos pueden estar correlacionados y son específicos de cada efecto.

A.6.3. Modelo de efectos aleatorios

Sea:

$$y_{ij} = a_i + b_j + e_{ij}$$

donde a y b son variables aleatorias, del mismo modo que el error e. Para cada aleatoria se define una variancia correspondiente: σ_a^2 ; σ_b^2 ; σ_e^2 .

La variancia de y será (asumiendo que las covarianzas entre las variables aleatorias son 0 (cero)):

$$V_y = \sigma_a^2 + \sigma_b^2 + \sigma_e^2 \neq I\sigma_e^2$$

A.6.4. Modelo lineal mixto

Sea:

$$y_{ij} = a_i + b_j + e_{ij}$$

donde a es un efecto fijo y b es una variable aleatoria, del mismo modo que lo es el error e . Para cada variable aleatoria se define su variancia: σ_b^2 ; σ_e^2 .

La variancia de y será V_y (asumiendo que las covarianzas entre las variables aleatorias son 0 (cero)):

$$V_y = \sigma_b^2 + \sigma_e^2 \neq I \sigma_e^2$$

Nos referiremos en esta sección a los modelos lineales donde las variables clasificatorias son de efectos fijos y de efectos aleatorios. El modelo expresado de manera matricial es el siguiente:

$$y = Pb + Zu + e$$

donde b , u y e son vectores de los efectos contemplados en el modelo. El vector e corresponde al error o al residual no considerado por el modelo, b corresponde a los efectos fijos y u es el vector de los efectos aleatorios, excepto el error experimental. Las matrices P y Z tienen elementos iguales a 0 o a 1, según que el efecto al que hacen referencia esté ausente o presente en el valor observado y .

Cuando consideramos un modelo lineal general (GLM), los efectos que se incorporan al modelo son todos fijos, con excepción del error experimental:

$$y = Pb + e'$$

donde $e' = Zu + e$ del modelo mixto (Walsh, 2012).

En el enfoque que nos interesa en este libro, la matriz Z tiene que ver con los efectos genéticos o aditivos de la muestra aleatoria de genotipos evaluados.

La variancia de y , de acuerdo al modelo será:

$$V(y) = Z'Var(u)Z + Var(e) = Z'GZ + I\sigma_e^2,$$

donde I es una matriz identidad y G es la matriz de correlaciones genéticas o entre valores aditivos entre individuos, es decir, que:

$$G = A \sigma_g^2$$

donde $A = 2r_{MN}$, donde r_{MN} es el coeficiente de coancestría entre dos individuos M y N (ver Capítulo 5). La matriz A es conocida por cuanto depende de los pedigrís.

Por lo tanto la variancia de los valores fenotípicos será:

$$Var(y) = Z'AZ \sigma_g^2 + I\sigma_e^2$$

Frecuentemente se designa:

$$R = I\sigma_e^2$$

Por lo tanto

$$Var(y) = V = Z'GZ + R$$



En los modelos fijos la tarea es estimar a los efectos fijos y predecir los efectos aleatorios. En el primer caso se estiman los BLUE (*Best linear unbiased estimators* o mejores estimadores lineales insesgados, o LSMEANS) y en el segundo se predicen los BLUP (*Best linear unbiased predictors* o mejores predictores lineales insesgados).

A.6.4.1. Predictores BLUP

Previo a obtener los BLUPs necesitamos estimar los componentes de variancia mediante el empleo de análisis de variancia, o aplicando los métodos de máxima verosimilitud o de máxima verosimilitud restringida.

Los mejores predictores lineales insesgados (BLUP) de mérito genético de los genotipos pueden obtenerse como solución a las ecuaciones de modelo mixto (EMM) propuestas por Henderson (1976).

Ello requiere la inversión de la matriz de parentesco G.

Las ecuaciones del modelo mixto son:

$$\begin{pmatrix} P'P & P'Z \\ Z'P & Z'Z + \sigma_e^2/\sigma_g^2 A^{-1} \end{pmatrix} \begin{pmatrix} b \\ u \end{pmatrix} = \begin{pmatrix} P'y \\ Z'y \end{pmatrix}$$

Las matrices inversas de R y G son, respectivamente:

$$R^{-1} = 1/(\sigma_e^2) * A^{-1}$$

$$G^{-1} = 1/(\sigma_g^2) * A^{-1}$$

La relación (σ_e^2/σ_g^2) puede reemplazarse en términos de heredabilidad en sentido estricto:

$$(\sigma_e^2/\sigma_g^2) = \lambda = (1-h^2)/h^2$$

A partir de los datos conocidos o deducibles (vector y, matrices de incidencia P y Z) el objetivo es obtener estimadores de los efectos fijos (vector b) y de los efectos aleatorios (vector u y e). Para ello necesitaremos estimar las matrices R y G, es decir, la de los residuales y de las correlaciones genéticas entre los individuos evaluados. Las estimaciones del vector b y los predictores del vector u serán:

$$b_e = (P'V^{-1}P)^{-1} P'V^{-1}y$$

$$u_e = GZ'V^{-1}(y - Pb_e)$$

La aplicación de las ecuaciones del modelo mixto permite obtener las estimaciones de máxima verosimilitud o REML de los efectos fijos y de los aleatorios.

Puede demostrarse que el predictor del valor de mejora del i-ésimo individuo es (Walsh, 2012):

$$BLUP(i) = h^2(y - \mu), \text{ donde } \mu \text{ es la media general.}$$

Para seleccionar el modelo correcto cuando el modelo mixto es anidado puede emplearse el criterio de la relación del logaritmo de verosimilitud. En consecuencia es necesario calcular el logaritmo de la verosimilitud REML de un modelo reducido (LL_{RMR}) y el logaritmo de la verosimilitud del modelo completo (LL_{RMC}) de tal modo que (Balzarini, 2002):

$$L = -2 (LL_{RMR} - LL_{RMC})$$

Si ambos modelos son igualmente eficientes, el valor $\chi^2_{(p)}$ no resultará significativo, para p grados de libertad definido por la diferencia de parámetros considerados en el modelo completo y el modelo reducido.

Cuando los modelos mixtos que se comparan no son anidados, la elección del modelo debe basarse en el criterio de información de Akaike (AIC).

A.7. Medidas de distancia

A.7.1. Distancia euclidiana

Es una de las medidas de distancia más frecuentemente utilizada ya sea cuando se comparan valores fenotípicos o genotípicos y medias ambientales.

Por ejemplo, la distancia euclidiana entre dos genotipos j y j' en el ambiente i -ésimo se define como:

$$d_{jj'}^2 = \sum_{i=1}^s (y_{ij} - y_{ij'})^2$$

donde y_{ij} e $y_{ij'}$ son los comportamientos de ambos genotipos en el ambiente común.

Cuando se consideran múltiples genotipos y ambientes, las distancias euclidianas pasan a ser los elementos de una matriz de distancias.

A.8. Homogeneidad de variancias

En el análisis de variancia de una serie de experimentos en forma conjunta resulta muy importante determinar si las variancias del error de cada uno de los experimentos son homogéneas entre sí. Una prueba comúnmente utilizada para determinarlo es la de Bartlett (1937). Debe compararse el valor de Chi cuadrado calculado con el tabulado para el nivel de probabilidad deseado y $k-1$ grados de libertad:

$$X^2 = (1/C) [gl_{eec} \ln CM_{eec} - \sum (gl_{eek} \ln CM_{eek})]$$

donde:

gl_{eec} es el número de grados de libertad del error experimental combinado a través de k experimentos,

CM_{eec} es el cuadrado medio del error experimental combinado,

gl_{eek} es el número de grados de libertad del error experimental del k -ésimo experimento individual,

CM_{eek} es el cuadrado medio del error experimental del k -ésimo experimento,

y

$$C = 1 + \{1/[3(k-1)]\} [\sum (1/gl_{eek}) - (1/gl_{eec})]$$

Si todos los experimentos de la serie tuvieran el mismo número de grados de libertad para la variancia del error experimental, las expresiones indicadas más arriba se simplificarían de la siguiente manera:

$$X^2 = (1/C) [gl_{ee}(k \ln CM_{eec} - \sum \ln CM_{eek})],$$

$$C = 1 + [(k + 1)/(3k gl_{ee})]$$



A.9. Producto de Kronecker entre matrices

Dadas dos matrices $A_{n \times m}$ y $B_{p \times q}$, su matriz producto de Kronecker, $C_{n \times p \times m \times q} = A \otimes B$, tendrá como elementos:

$$c = (a_{ij} B)_{ij}$$

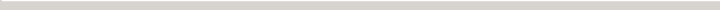
Si por ejemplo:

$$A_{3 \times 3} = \begin{pmatrix} 4 & 1 & 40 \\ 11 & 1 & 18 \\ 7 & 5 & 14 \end{pmatrix} \quad y \quad B_{2 \times 2} = \begin{pmatrix} 15 & 21 \\ 11 & 5 \end{pmatrix}$$

entonces:

$$C_{6 \times 6} = \begin{pmatrix} \left[\begin{matrix} (4 \times 15) & (4 \times 21) \\ (4 \times 11) & (4 \times 5) \end{matrix} \right] \left[\begin{matrix} (1 \times 15) & (1 \times 21) \\ (1 \times 11) & (1 \times 5) \end{matrix} \right] \left[\begin{matrix} (40 \times 15) & (40 \times 21) \\ (40 \times 11) & (40 \times 5) \end{matrix} \right] \\ \dots \\ \left[\begin{matrix} (7 \times 15) & (7 \times 21) \\ (7 \times 11) & (7 \times 5) \end{matrix} \right] \left[\begin{matrix} (5 \times 15) & (5 \times 21) \\ (5 \times 11) & (5 \times 5) \end{matrix} \right] \left[\begin{matrix} (14 \times 15) & (14 \times 21) \\ (14 \times 11) & (14 \times 5) \end{matrix} \right] \end{pmatrix}$$

$$C_{6 \times 6} = \begin{pmatrix} 60 & 84 & 15 & 21 & 600 & 840 \\ 44 & 20 & 11 & 5 & 440 & 200 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 105 & 147 & 75 & 105 & 210 & 294 \\ 77 & 35 & 55 & 25 & 154 & 70 \end{pmatrix}$$





Anexo - Índice

A.1. Medidas de tendencia central	461
A.1.1. Media aritmética o promedio aritmético	461
A.1.2. Otras medidas de tendencia central	462
A.1.2.1. <i>Media geométrica (G)</i>	462
A.1.2.2. <i>Media armónica (H)</i>	462
A.2. Medidas de dispersión	462
A.2.1. Variancia	462
A.2.2. Coeficiente de variación	463
A.3. Medidas de asociación entre variables	464
A.3.1. Regresión	464
A.3.2. Correlación	465
A.3.3. Relación entre regresión y correlación simple	465
A.3.4. Regresión lineal múltiple	465
A.3.5. Correlación parcial	467
A.3.6. Correlación múltiple	467
A.3.7. Regresión utilizando variables estandarizadas	467
A.4. Modelos lineales	469
A.4.1. Medias de mínimos cuadrados (LSMeans)	470
A.4.2. Variancia de funciones lineales	471
A.4.3. Covariancia de funciones lineales	472
A.5. Estructura de tratamientos	472
A.6. Esperanzas de los cuadrados medios en el análisis de variancia	473
A.6.1. Reglas de Schultz	473
A.6.2. Modelo de efectos fijos	477
A.6.3. Modelo de efectos aleatorios	477
A.6.4. Modelo lineal mixto	478
A.6.4.1. <i>Predictores BLUP</i>	479
A.7. Medidas de distancia	480
A.7.1. <i>Distancia euclidiana</i>	480
A.8. Homogeneidad de variancias	480
A.9. Producto de Kronecker entre matrices	481

El conocimiento y utilización de la variabilidad genética, expresable tanto con el auxilio de marcadores moleculares como de manera convencional por observaciones en individuos emparentados, continúa siendo la base para la toma de decisiones estratégicas en los programas de fitomejoramiento. Esta obra aborda los principios de la Genética Cuantitativa que guían esas decisiones en las diferentes etapas del mejoramiento, cualquiera sea su enfoque, de modo de contribuir a una mayor eficiencia en el uso de recursos físicos y por unidad de tiempo. Abarca consideraciones sobre el muestreo en las poblaciones de crianza, la cuantificación de la heredabilidad, la predicción de las respuestas a la selección para comparar diferentes estrategias selectivas, la selección genómica y la fenotípica, la selección simultánea para varios caracteres de interés, cómo aproximar la respuesta real al potencial alcanzable, y la amplitud y características que debe tener un programa de evaluación eficiente. Estos temas y otros directamente relacionados se tratan a lo largo de 14 capítulos y un Anexo, todos ellos acompañados por tablas y figuras explicativas.

ISBN 978-987-679-336-0



Ministerio de Agricultura,
Ganadería y Pesca
Argentina